NAVY EXPERIMENTAL DIVING UNIT

REPORT NO. 2-99

STATISTICALLY BASED CO_2 CANISTER DURATION LIMITS FOR CLOSED-CIRCUIT UNDERWATER BREATHING APPARATUS

J.R. CLARKE

NAVY EXPERIMENTAL DIVING UNIT

DEPARTMENT OF THE NAVY
NAVY EXPERIMENTAL DIVING UNIT
321 BULLFINCH ROAD
PANAMA CITY, FLORIDA 32407-7015

IN REPLY REFER TO:
NAVSEA TA 97-016

NAVY EXPERIMENTAL DIVING UNIT

REPORT NO. 2-99

STATISTICALLY BASED CO_2 CANISTER DURATION LIMITS FOR CLOSED-CIRCUIT UNDERWATER BREATHING APPARATUS

J.R. CLARKE

APRIL 1999

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

Submitted:
J.R. CLARKE
GM-15
Scientific Director

Reviewed:
R.L. JOHNSON
LCDR, USN
Senior Projects Officer

J.R. O'ROURKE
LCDR, USN
Executive Officer

M.D. CURLEY
CAPT, M.S.C.
Research Department Head

Approved:
E.N. CHRISTENSEN
CDR, USN
Commanding Officer

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT DISTRIBUTION STATEMENT A: Approved for Public Release
2b. DECLASSIFICATION/DOWNGRADING AUTHORITY	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NEDU TR No. 2-99	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Navy Experimental Diving Unit / 6b. OFFICE SYMBOL (If Applicable) 00D	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) 321 Bullfinch Road, Panama City, FL 32407-7015	7b. ADDRESS (City, State, and Zip Code)
8a. NAME OF FUNDING SPONSORING ORGANIZATION Naval Sea Systems Command / 8b. OFFICE SYMBOL (If Applicable) 00C	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) 2531 Jefferson Davis Highway, Arlington, VA 22242-5160	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. 96-014 / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
(U) STATISTCALLY BASED CO_2 CANISTER DURATION LIMITS FOR CLOSED-CIRCUIT UNDERWATER BREATHING APPARATUS

12. PERSONAL AUTHOR(S)
J.R. Clarke

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM	1999, APR	45

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
carbon dioxide absorbent, sodalime, curve fitting, , probability, confidence limits, prediction limits, canister durations, rebreathers, closed-circuit UBA

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
Elementary statistical methods are used to predict canister durations as a function of gas temperature and CO_2 production rate. The canister duration data is fit to a 3 term polynomial using TableCurve 2D or similar curve fitting software. The best fit curve and the standard error of the estimate are used to derive prediction limits of varying degrees of confidence. Examples are given of the use of prediction limits in biomedical and manufacturing applications. From the prediction limits we can define a series of canister duration curves, with increasing duration being associated with an increasing risk of canister breakthrough.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED X SAME AS RPT. DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
NEDU Librarian	904-230-3100	

DD Form 1473

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

CONTENTS

ILLUSTRATIONS

TABLES

GLOSSARY

Canister duration	The elapsed time from the beginning of either real or simulated breathing until the partial pressure of CO_2 in the canister effluent consistently exceeds 0.5 kPa (0.5% SEV).
Canister limit	The working time allowed for a canister under particular environmental and work conditions. It is generated by applying a statistically appropriate safety factor to the average canister duration for a given environmental condition.
Confidence limits (on the mean)	A measure of the statistical certainty of population means. Confidence limits comprise the upper and lower boundaries of a confidence interval, which define a range of values in which there is a specific likelihood (level of confidence) that a parameter (such as a mean) will fall.
FAA	Federal Aviation Administration
Prediction limits	A measure of the statistical certainty of future measurements made under identical conditions. An upper and a lower limit define the prediction interval expressing some level of confidence in the appearance of future measurements. Canister limits are equal to the lower of the prediction limits.
NAVSEA	Naval Sea Systems Command
NEDU	Navy Experimental Diving Unit
SEV	Surface equivalent value. One way of expressing the partial pressure of a gas referred to 1 ATA. The preferred way of expressing partial pressure is in pressure units, i.e. kPa. A 0.5% SEV is approximately equal to 0.5 kPa.
STPD	Standard temperature (0°C), pressure (1 atm abs), dry. 760 mmHg, with 0 mmHg water vapor pressure.
$\dot{V}CO_2$	CO_2 production expressed in $L \cdot min^{-1}$ at STPD conditions.
$\dot{V}O_2$	Metabolic oxygen consumption in $L \cdot min^{-1}$ at STPD conditions.

INTRODUCTION

For the past several years, NEDU has been applying statistical techniques used in medicine and industry to determine how long a diver can safely use a CO_2 absorbent canister in a closed or semi-closed circuit underwater breathing apparatus (UBA). Those UBA are often referred to as rebreathers. The current statistical techniques differ from the methods used at NEDU ten to twenty years ago, and have been described in part in several classified NEDU technical reports. The purpose of this report is to explain the rationale for our current statistical practices, and to expand upon our descriptions of the methodology. Our approach will be intuitive and largely non-mathematical.

Why Statistics?

A quote originating from a researcher at the Memorial Sloan-Kettering Cancer Center[1] and used in the Primer of Biostatistics[2] reflects the essence of this report.

> Hunches and intuitive impressions are essential for getting the work started, but it is only through the quality of the numbers at the end that the *truth* (current author's italics) can be told.[1]

Throughout this report we will be concerned with statistical methods for revealing the truth.

Twenty-five years ago, the U.S. Navy was seeking a CO_2 monitor for use in rebreathers[3]. Unfortunately, the engineering difficulties have been daunting; the search for such a monitor continues to this day. Consequently, when NEDU determines the length of time that a CO_2 absorbent canister can be safely used in UBA, we measure the time required for CO_2 to start passing all the way through the canister during simulated dives. CO_2 leakage, called canister break-through, is caused by partial depletion of the CO_2 absorption ability of the canister. From the few measurements made in our laboratory, we must infer how long the canisters will protect divers from CO_2 poisoning when used by the fleet. The process of inferring the outcome of 1000+ real dives based on the results of 20 simulated dives is one that requires carefully thought out statistical techniques. Those techniques are described below.

THEORY

Definitions

NEDU has historically defined a canister *duration* as the elapsed time from the beginning of either real or simulated breathing until the partial pressure of CO_2 in the canister effluent consistently exceeds 0.5 kPa (0.5% relative to 1 ATA or surface equivalent value (SEV)). The canister duration is strongly influenced by absorbent characteristics, canister design, absorbent bed moisture, pressure, temperature, and diver work rate. On the other hand, a canister *limit* is the working time allowed for a canister under particular environmental and work conditions. It is generated by applying a statistically appropriate safety factor to the average canister *duration* for a given environmental condition.

In a perfect world

For reasons of conjecture, we will assume that we in fact have the technology to monitor CO_2 levels in a UBA, and that the monitor is reliable; as reliable as the pressure gauge in SCUBA, or as the fuel gauges in jet aircraft. With such CO_2 monitors it should be theoretically possible to continue a dive until the monitor reads "E" for empty and then to surface. For a diver that would, of course, be unwise. Cave divers and pilots both apply safety margins to their operations to account for unexpected events. Those safety margins have been built up through the years based on experience – which is to say that as safety data is accumulated, operating policies have been established based upon a statistical review of the available data, tempered with carefully conducted risk/benefit analyses.

As an example, cave divers operate on the 1/3 rule. Two divers will continue a cave penetration until 1/3 of their gas is expended, then turn around. If one of the buddy pair has a total gas failure, then there should be enough gas remaining in the other diver's bottle to get both of them back to the surface. That gas supply safety factor has been found adequate to cover most contingencies.

Of course, cave divers still run out of gas and drown, but the basic policy remains unchanged because of the cost/benefit analysis that says carrying even more on board gas becomes a logistical and mobility burden, for very little potential payoff. In other words, the probability that a diver's life will be saved by expanding his percentage of extra gas (a clear benefit) becomes very low compared to the risk that the extra gas will compromise or limit the diver's mission. The application of statistics and probability that goes into these consensus opinions within cave diving training organizations may be informal, but they never-the-less occur.

Airlines and the FAA have likewise required operational safety margins for fuel management. Every pound of fuel carried on board means one less pound of paying passenger. But running out of fuel in flight is an event that airlines and the FAA strenuously avoid. So how are the risk/benefits managed?

FAA regulations require that an aircraft operating on an Instrument Flight Plan must carry enough fuel on board to reach their destination, attempt an approach, then if necessary fly to an alternate destination known to have better weather, and still land with 45 min of fuel on board. This rule keeps the majority of aircraft out of the "fuel exhaustion" accident category. But not always; even commercial aircraft have run out of gas in route due to unexpected circumstances.

In summary, even with accurate, certified instrumentation to indicate the supply of a critical expendable substance like air or fuel, and with safety procedures and margins developed from accident statistics and cost/benefit analyses, unforeseen events still cause people to die.

Diving Reality

In a rebreather, there are two expendables that sustain a diver's life as long as they are available; namely oxygen and CO_2 scrubbing ability. Unfortunately, only one of these can be measured – the O_2 supply. Lack of a monitor of CO_2 scrubbing effectiveness would not be a problem as long as it could be assured that the scrubber would always outlast the oxygen supply. If the critical expendable can be monitored, then the more abundant expendable can be ignored.

In fact, however, scrubber effectiveness depends on a multitude of factors, including water temperature, moisture content of the bed, absorbent quality, flow rate through the absorbent canister, the packing of the absorbent bed, and the CO_2 production rate of the diver. The latter depends in turn on the diver's work rate and oxygen consumption rate, and a proportionality "constant" called R, the respiratory exchange ratio. That ratio determines how much CO_2 is produced for a given amount of O_2 consumed. An R of 0.8 means that 0.8 moles of CO_2 are produced for each mole of oxygen consumed metabolically. Unfortunately, R is not really a constant – it normally varies between 0.7 and 1, depending upon the diver's diet, workload, and even temperature[4]. During strenuous exertion[5] it can climb as high as 1.5.

The result of all these confounding factors is that we can never be sure that a CO_2 canister's scrubbing ability will outlast the diver's monitored oxygen supply. The situation is somewhat akin to hoping that the capacity of a commercial aircraft's sewage holding tank outlasts the plane's fuel supply. It usually does, but passengers that frequent the rear of aircraft occasionally notice the distinctive odor indicating a mismatch between

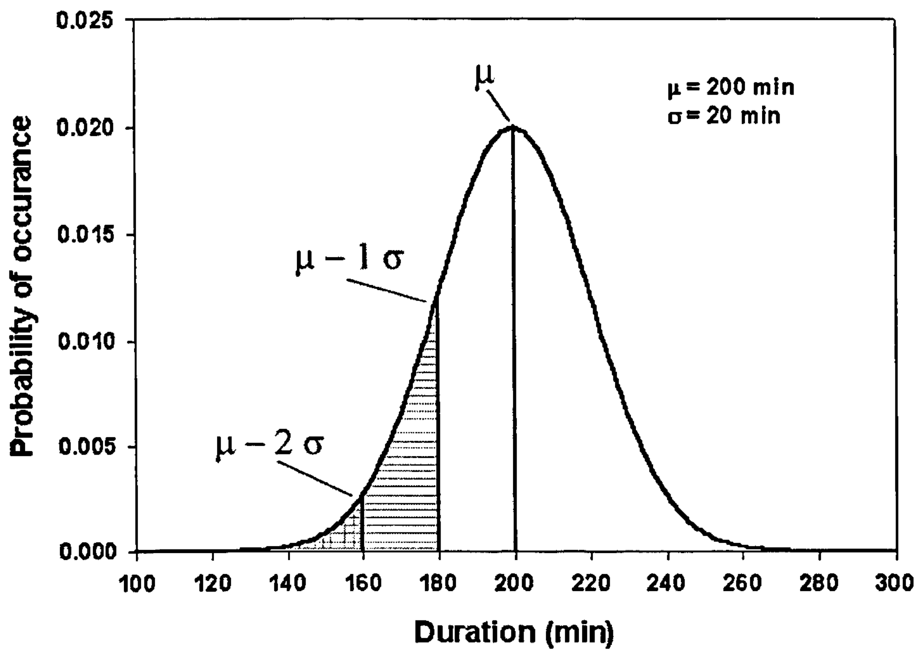


Figure 1. A normal (Gaussian) distribution. See text for details.

the ideal situation and reality. The big difference between the airline passenger and a diver is that when a diver's gaseous "sewage tank" overflows, the diver becomes unconscious.

Safety Factors

The key to keeping a diver alive amidst all the unpredictable confounding factors described above is the judicious use of statistically based canister duration limits. What does "statistically based" mean? It can be nothing more than measuring the average or "mean" time that a canister will last before overflowing. The mean (μ) is a statistic. Going one step further, one can also measure how the canister durations vary around that mean. A common measure of that variability is called the standard deviation (σ), the magnitude of deviation of canister endurance around the mean duration. A large σ means that some canisters last much shorter or longer than the average. A small σ means the canister durations are tightly clustered around the average duration.

A statistically based canister duration could simply be a duration based upon the mean duration. By definition of the mean or average, in a large population of canisters half of the canisters will last less than the average, and approximately half will last longer. Thus a canister limit based upon the mean statistic is one that allows half of CO_2 canisters to exceed the prescribed CO_2 limit. This event is called canister breakthrough.

The next step up in conservatism is to subtract one σ from the mean, and establish that as a canister duration limit. In theory, if one conducted very many tests (> 1,000) of canister duration under identical conditions, then subtracting one σ from the mean would provide a duration that only 16% of the canisters would breakthrough prior to reaching. Subtracting 2 σs from the mean yields a limit exceeded by only 2.3% of canisters.

These points are illustrated in Figure 1 which shows the probability density function for a normal or Gaussian distribution with a mean of 200 min, and a standard deviation of 20 min. The y axis gives the probability of finding a particular canister duration. The area underneath the curve represents cumulative probability and sums to 1.0. The probability of finding a canister duration less than some duration X is equal to the area underneath the curve to the left of X. For instance, the probability of finding a canister duration less than 300 min is essentially 1.0. The probability of encountering a canister duration less than 200 min is one half of 1.0 or 0.5. That is, half of the durations (area in grey including striped and cross-hatched areas) will be shorter than the mean.

All of the area to the left of the 180 min line (mean minus one σ, horizontal line shading) represents 16% of the total area, thus the probability of finding a duration equal to or less than 180 min is 16%. The area to the left of the 160 min line (mean minus two σ, cross-hatched area) occupies 2.3% of the total. Thus the risk of encountering durations less than or equal to 160 min is 2.3% or less.

Probability Calculations

Anyone who wants to confirm the above percentages can download from the Internet a free Windows-based probability calculator. It is located at the following WEB address:

http://www.ncss.com/download.html

Once the program is running, the normal probability distribution (one of many) can be selected. Under the "Input" column, the distribution mean and sigma (σ) is entered, followed by the canister duration (X) of interest. The probability of encountering a duration less than or equal to X is then calculated and displayed under the column labeled "Normal Results."

In general, knowing the mean (μ) and standard deviation (σ) of a large sample of canister durations, one can determine the duration ($T_{\%risk}$) that will yield a risk of breakthrough equal to the top row of Table 1 by subtracting from the mean a value equal to $K \cdot \sigma$, where K is given below its respective % risk

$$T_{\% risk} = \mu - K_{\% risk} \cdot \sigma \tag{1}$$

Table 1. % risk and σ multipliers.

% risk	2.5	5	10	15	20	25	30	40	50
K	1.960	1.645	1.282	1.036	0.842	0.674	0.524	0.253	0

To illustrate this, we used a computer to simulate 5,000 canister durations distributed in a normal or Gaussian manner. (Early versions of the commercial software package *StatGraphics* (STSC, Rockville, MD) and recent versions of *MathCad* (MathSoft, Cambridge, MA) can produce pseudo-random numbers distributed in a variety of distribution shapes.) The measured mean for the population was 200 min, with a σ of 20 min, just as in Figure 1. By using equation 1 and Table 1, we obtain a table of canister duration limits that provide increasing risk of canister breakthrough (Table 2). By sorting the 5,000 "durations," we were able to confirm that indeed only 125 durations (2.5%) were shorter than 161 min, and that 500 durations (10%) were less than 174 min.

We conclude that when dealing with very large data sets (thousands of data points) that canister duration and the risk of canister breakthrough can be accurately predicted by subtracting some fraction of the standard deviation from the mean canister duration. Furthermore, the greater the tolerated risk, the longer the canister duration limit.

Table 2. % risk and duration limits.

% risk	2.5	5	10	15	20	25	30	40	50
T(min)	161	167	174	179	183	186	189	195	200

Sampling Error

Cost, time, and manpower constraints prevent NEDU from performing a large number of canister duration runs under a single environmental condition. We are forced to make inferences about canister durations based upon a small number of tests. Unfortunately, small sample sizes do not estimate the total population of canister durations particularly well, leading to potentially large errors in our estimates.

To explore the implications of sampling errors, we again used the above mentioned large population of simulated canister durations. We will refer to that population as the "Truth" data set. When plotting the frequency of occurrence of various canister durations, the Truth data set (Figure 1) yielded the familiar bell shaped curve centered upon a duration of 200 min, and with 68.3 % of the canister durations falling within 19.8 min on either side of the mean.

At NEDU we typically perform five canister duration runs in any rebreather under as close to identical conditions as possible. For the purposes of this report we can simulate that process by having a computer draw five samples randomly from the Truth data set, and then calculating the mean and standard deviation of those samples. Those sample statistics are our best estimates of the characteristics of the Truth data set. We can then apply some safety factor to those statistics to derive our published canister limit for the tested conditions. As stated in the previous section, if we subtract one sample σ from the sample mean, then we hope to derive a canister limit that produces only a 16% risk of canister breakthrough if the canister is dived to the limit.

In Figure 2, we see that our goal is elusive. The solid circles are the canister duration limits which resulted from 10 trials of five samples each of the Truth data set. *Remember, in reality we only get one trial.* The horizontal lines are the canister limits that would actually result in a breakthrough risk indicated at the right extreme of each line. Although we expected a duration limit that would result in only a 16% chance of canister breakthrough, what we actually received varied between a 4% and 49% risk.
In other words, based on any one trial we have virtually no confidence in our risk assessment for the canister duration limit.

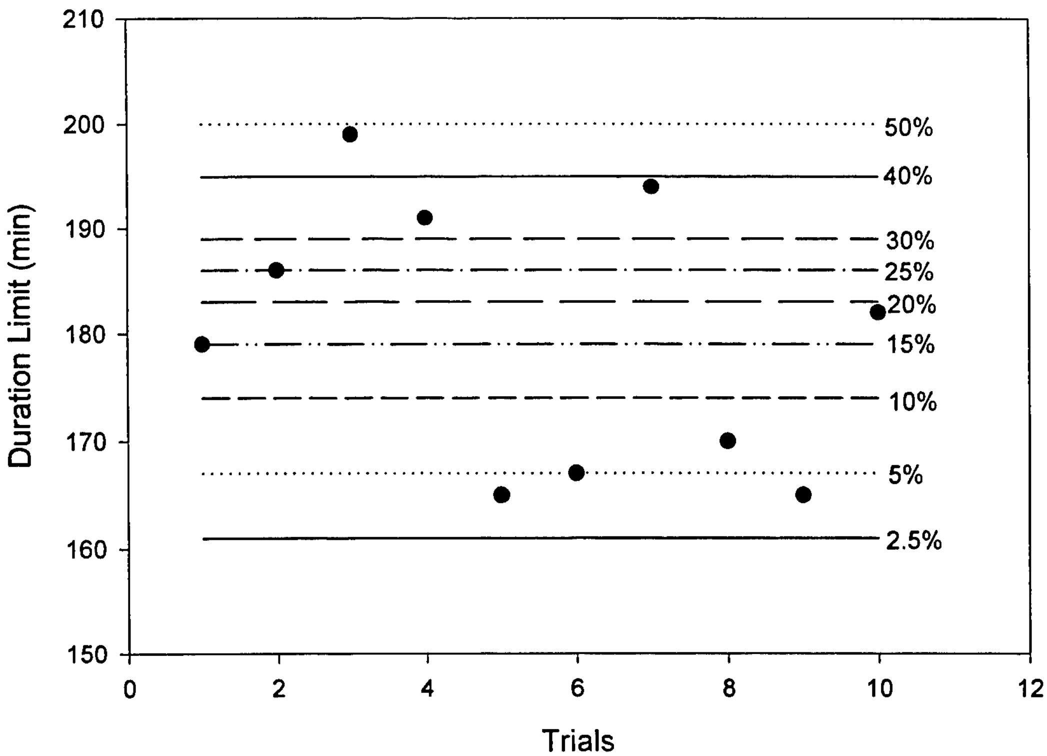


Figure 2. Multiple trial samplings from a population of canister durations. Canister duration limit for each sample was found by subtracting the sample standard deviation (σ) from each sample mean (μ).

A Range of Temperatures

The above demonstration was for multiple trials at a single environmental condition. Canister durations tend to decrease at deeper depths, and can decrease dramatically at low water temperatures. Typically, NEDU determines canister durations over a range of depths, and with temperatures ranging from 28°F to 90°F. At each environmental condition we still take five samples as described above. (Although five samples is our goal, technical difficulties occasionally limit us to as few as three tests per environmental condition.).

In Figure 3 we plotted the results of repetitive trials, five samples per trial, one trial for each of five water temperatures. Mean durations are indicated by solid circles, the old NEDU limit of mean minus σ is shown as open circles. Since these trials are computer generated we have the luxury of repeating each test as often as we wish. Here we have plotted the results of four such repetitions. To simplify matters for this

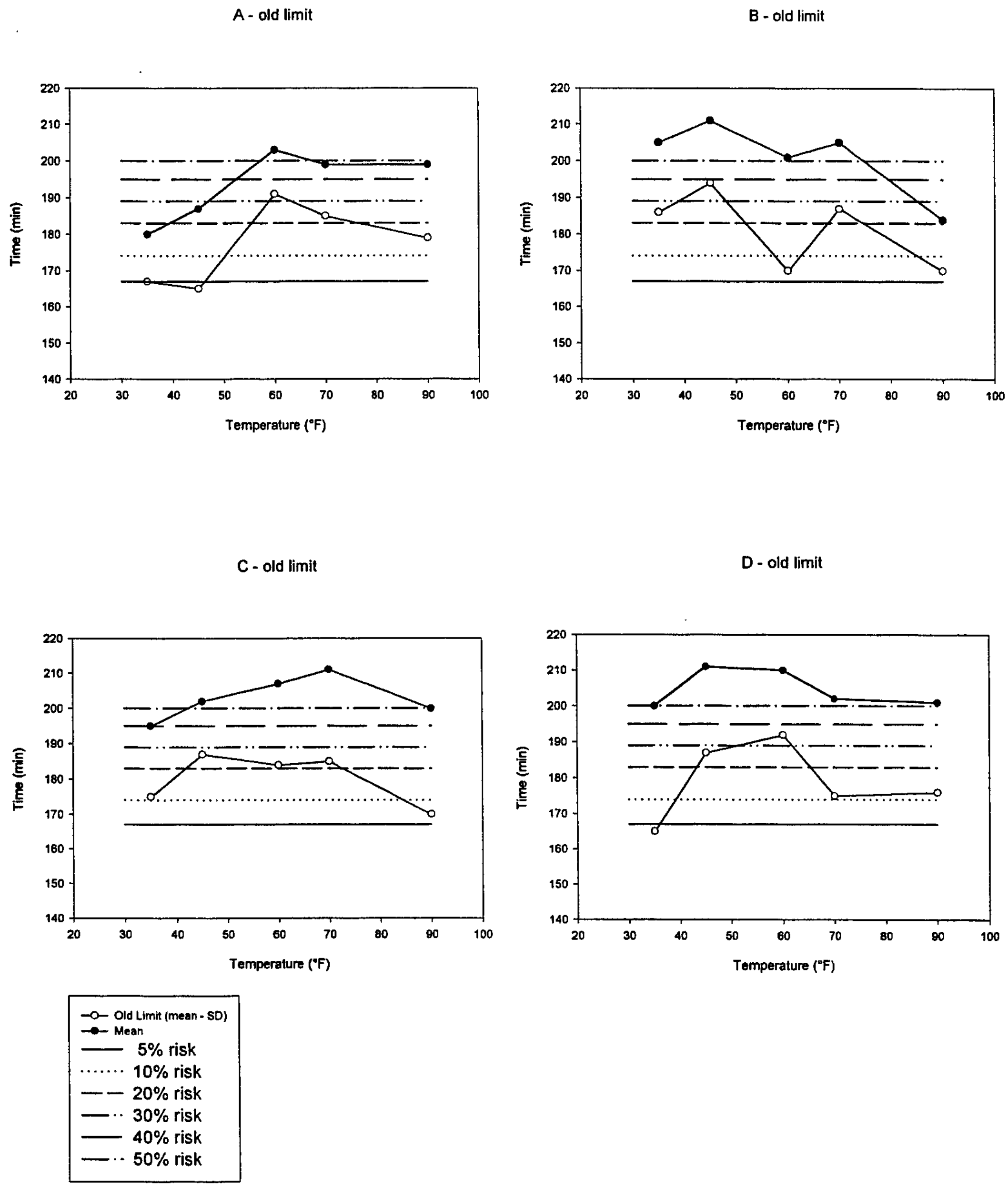


Figure 3. Multiple samples across a range of temperatures. Filled circles are sample means, and open circles are the limits derived for each temperature by subtracting one sample σ from the sample mean. True risk of each limit is indicated by horizontal lines.

demonstration, we assumed that temperature had no effect on canister duration. All variation in mean durations and the resulting duration limits are attributable to sampling errors due to small sample sizes. Once again, the true risk of each canister limit, indicated by a horizontal line, varied greatly, from less than 5% to almost 40%.

Modeling

In the above example, we made no judgment as to whether or not measurements of canister durations taken at one temperature were related to similar measurements of duration at another temperature. Each sample was assumed to be independent of the others, each with its own mean and σ, and each with its own highly variable canister limit based upon the sample mean minus σ. From basic physico-chemical principles, however, we know that the measurements at one temperature are related to those at other temperatures. The relationship between them can be expressed by a mathematical equation which we call a "model."

Any model contains a mixture of known variables (such as temperature and pressure) and unknown parameters. Although initially unknown, those parameters can be estimated through a procedure called variously regression, curve fitting or parameter estimation.

Scientists are often interested in particular equations or models that have been suggested from first principles. There may be specific reasons why a particular model should apply to a given physical or chemical process, such as Newton's Second Law applying to a body in free fall. However, there are also situations when no particular model is preferred *a priori*. Any model that fits well to the data can be used to describe the data. That is indeed the situation in which we find ourselves when modeling canister duration data. We can use a computer to evaluate a multitude of potential models and find the one which best fits the data. The model that best fits our canister duration data is called an empirical model of canister performance.

The benefit of seeking an empirical model of canister performance is that the model dramatically reduces the number of parameters that must be estimated from the available data. For instance, in the examples above we assumed no relationship between measurements made at different water temperatures. We tested at five temperatures, and for each temperature estimated two parameters or statistics, namely the mean and standard deviation for each measurement. A total of 10 parameters were estimated based on a total of 25 canister durations. Because we have so many parameters to estimate from so few data points, we can not be very certain about any given estimate. That uncertainty results in the highly variable predictions seen in Figures 2 and 3.

The Value of Regression

We can make much better use of canister duration data by describing them with an empirical model. Such models are typically obtained through regression techniques, of which the method of least squares is one of the more ubiquitous. The successful prediction of an expected value of Y (for instance, canister duration) for a given value of X (temperature) is a major aspect of regression analysis, "and has been suggested as one justification for employing empirically fitted curves"[6]. In fact, "most regression lines are empirically fitted curves, in which the functions simply represent the best mathematical fit (by a criterion such as least squares) to an observed set of data"[6].

A regression-based model derived from fits to the canister duration data may have as few as two parameters if the durations are linear with temperature. Without the model, we would be forced to estimate five times as many parameters from the same data. The result of using the model is an increase in the confidence we have in making predictions about future canister durations. We'll see an example of that in a later section.

Prediction Limits in Regression

The use of prediction intervals in regression is an elementary statistical technique, particularly well adapted to making, as the name suggests, predictions about future but closely related events. Prediction intervals are introduced in virtually every beginning statistics text under the subject of regression, although the exact terminology varies. For instance, in Primer of Biostatistics[2] by S.A. Glanz (1992), the subject is referred to as "the confidence interval for an observation." In J.T. McClave and P.G. Benson, A First Course in Business Statistics[7] (1992), it is referred to as a "prediction interval for individual values."

More rigorous treatments are provided in Neter, Wasserman, and Kutner, Applied Linear Statistical Models[8] (1990) where in section 3.5, Prediction of New Observations, the authors refer to "prediction limits" and "prediction intervals." Likewise, in Sokol and Rohlf, Biometry[6], (1995), the subject is introduced as prediction limits.

Unfortunately, even among the more complete texts, there is a lack of consistency. In Cohen and Cohen, Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences[9], the topic is covered under "Confidence limits on a single Y value.". In Draper and Smith, Applied Regression Analysis[10], 2^{nd} ed. (1981), the subject is addressed as "confidence limits for new Ys." In general, modern statistical texts, either basic or advanced in approach, tend to emphasize the wording "prediction intervals and limits," and so shall we.

NEDU first used regression techniques and confidence intervals to develop canister limits for the NEDU report TR 2-93, *MK 16 Canister Limits for SDV Operations*[11]. Prediction intervals made their first appearance in NEDU TR 09-97, *Recommended Canister Limits for the Draeger LAR V/MK 25 UBA Using 408 L-grade and 812 D-Grade Sofnolime*[12].

The following three figures graphically demonstrate the meaning of prediction intervals. We used a computer to generate 1,000 data points selected in the following manner. An X value, representing temperature was chosen at random within the temperature range of 28° F to 128° F. A Y value corresponding to the selected X was then determined based on a known linear equation (for Figures 4 and 5) and a curvilinear equation for Figure 6. To each Y value computed from its governing equation an error term was added, with the sign of the error being chosen at random, and with the magnitude of the error being distributed in an approximately Gaussian manner. In other words, there was a high probability that an error term would be small, and that a selected Y value would lie close to the value predicted by its governing equation. The odds that an error term would be large, and that the "measured" Y would be located far from the mean location, was small. However, with enough repetitive data gathering, even low probability events do occur. It is immediately apparent, however, that the further a point lies from the ideal location, the fewer neighbors it has because its appearance is relatively unlikely.

Through a regression technique described in most elementary statistics texts, we computed lines that represent the best estimate of the correct Y for a given X, and two lines representing prediction limits. (Details of this process will be given in a following section of this report.) The estimate for the best Y is one that minimizes the impact of the error term for each data point, and which reproduces the dependence of Y on X defined in the governing equation. It is the same line that is commonly found by linear or nonlinear regression methods, and defines the average Y for a given X.

The lines representing a 95% prediction interval (Figure 4) lie at the boundaries of a region that contains approximately 95% of the measured data points. We thus infer that in future tests run under identical conditions, 95% of the new data will fall between those lines. We also conclude that half of the remainder, or 2.5%, will fall below the 95%

prediction limit, and 2.5% will occur above the upper prediction limit. When dealing with small sample sizes it is common to see no data points falling below the lower limit; their occurrence is relatively rare. However, with 1,000 computer generated data points, approximately 25 points should fall both above and below the 95% limit lines.

Figure 5 shows the consequences of decreasing our confidence in the prediction interval from 95% to 90%. When predicting new test results, only 90% will likely fall within the 90% prediction limits. Half of the remainder, or 5% will fall below the prediction line, and another 5% will reach above the upper prediction line.

The linear graphs of Figures 4 and 5 are similar to ones obtained during testing at NEDU of semi-closed UBA. The curvilinear graph in this series (Figure 6) shows that these predictions are not restricted only to linear conditions. The shape of Figure 6 is similar to that found during testing of fully closed UBA such as the MK 16 and the LAR V. In addition to prediction limits, Figure 6 also includes confidence limits about the mean. With those confidence limits, we can be 95% certain that in future tests the mean value of any Y for a given X will fall within the interval between those confidence limits. Due to the large number of data points, the confidence regions about the mean are narrow. As the number of data points decrease, our confidence in the estimates of the mean will decrease, leading to wider confidence limits.

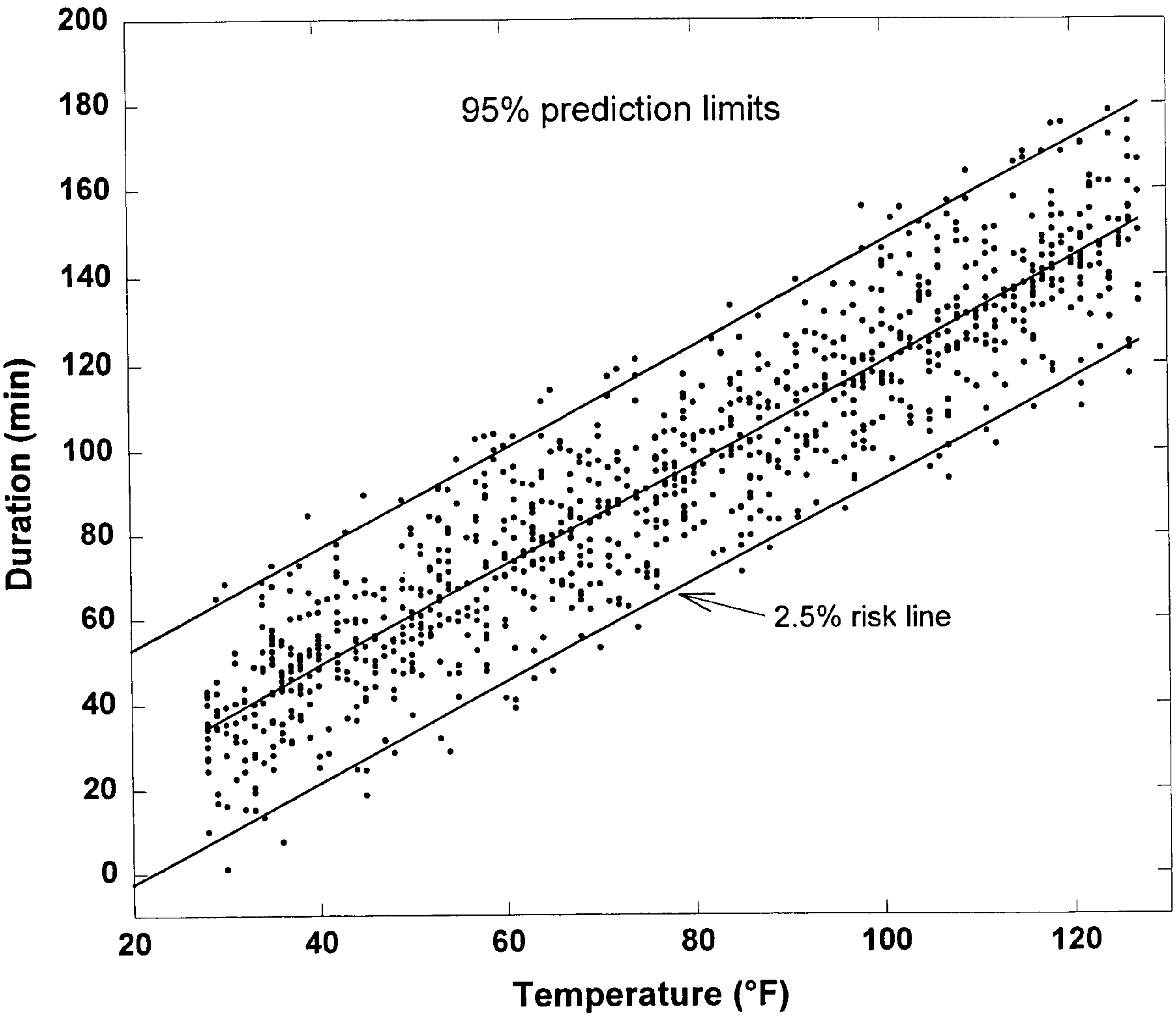


Figure 4. Linear regression and 95% prediction limits for 1,000 data points. The lower prediction limit defines a boundary below which only 2.5% of canisters are expected to fall.

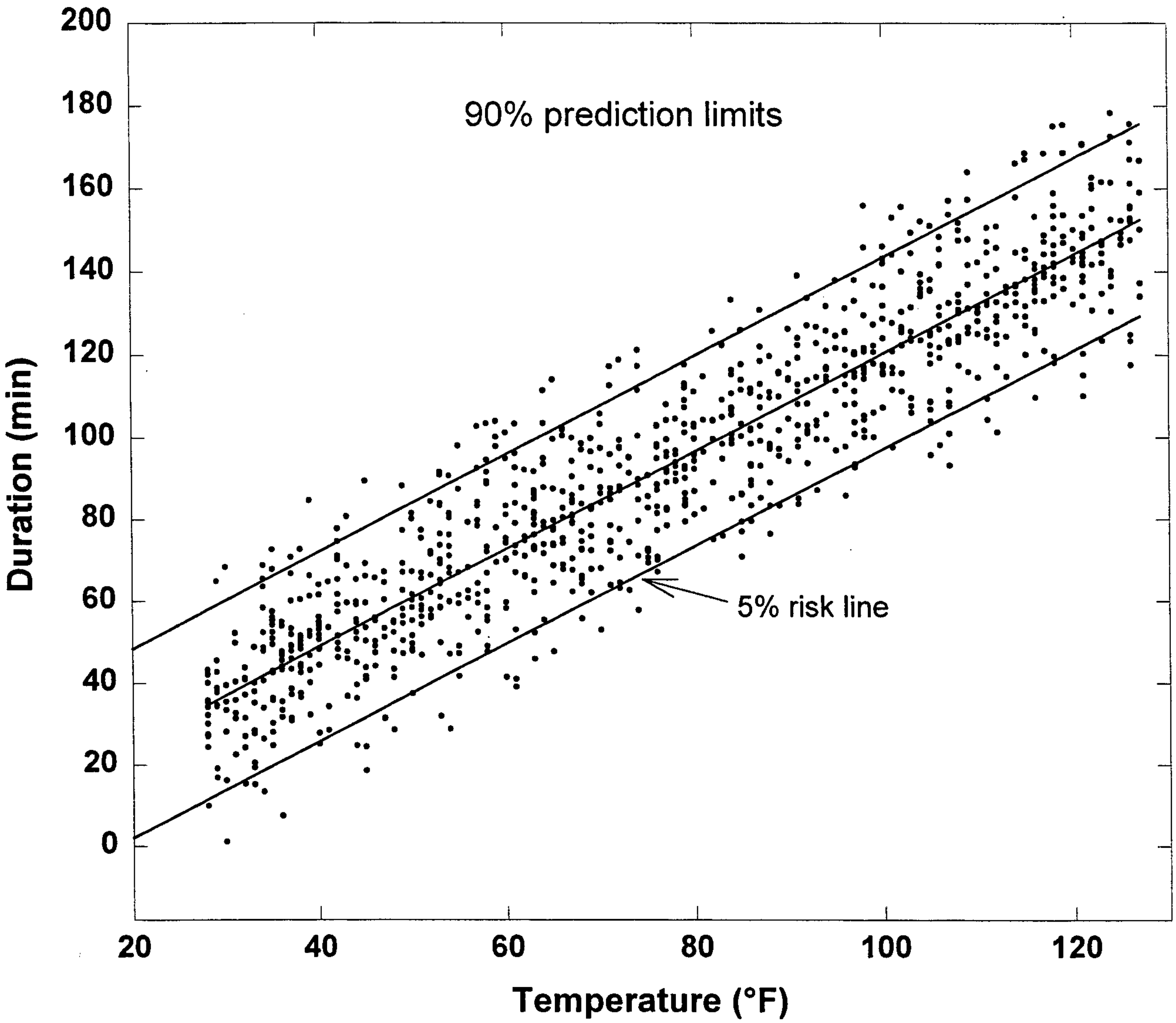


Figure 5. 90% prediction limits on the regression.

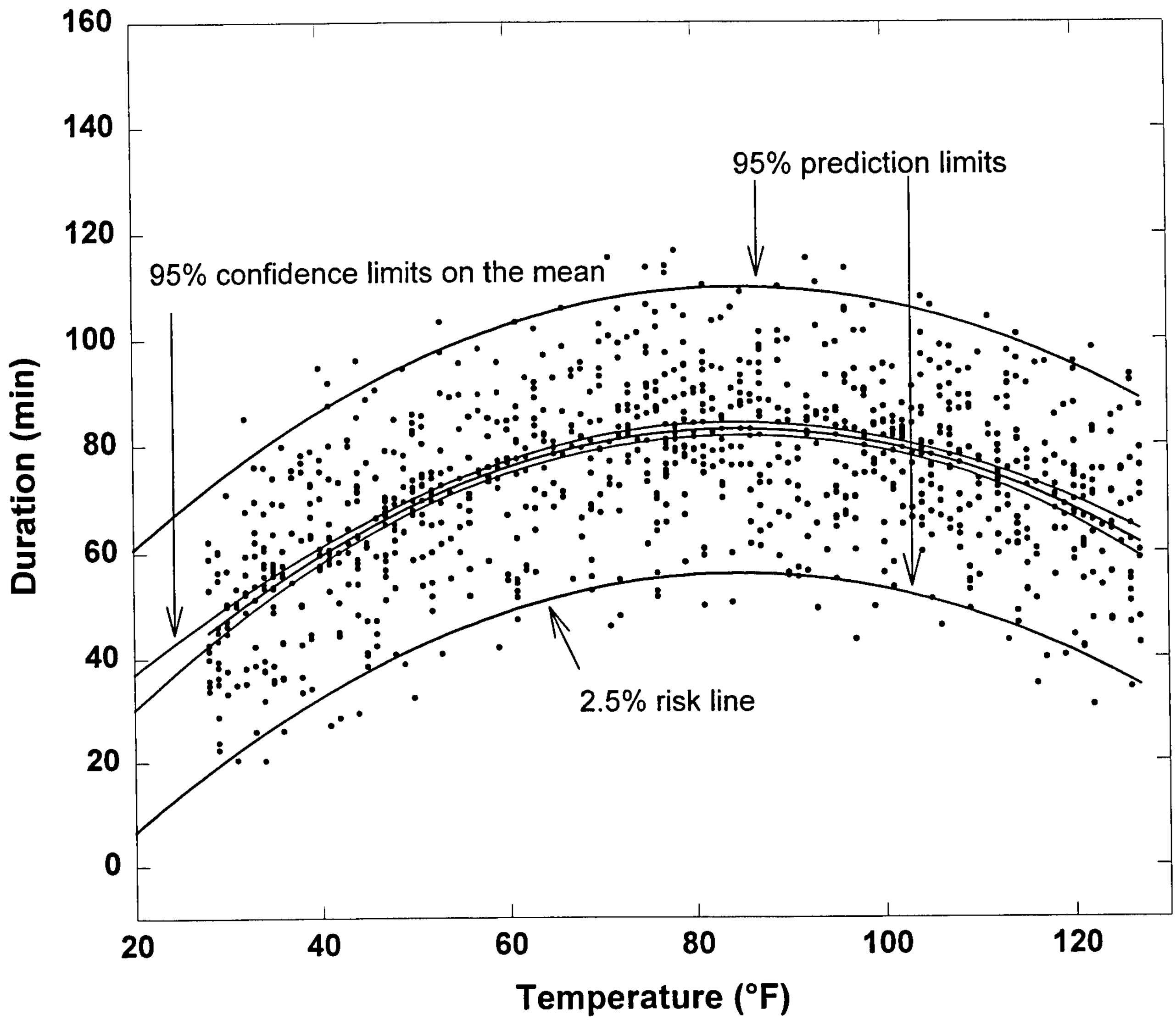


Figure 6. Regression of simulated durations nonlinearly dependent on temperature. Curves show the 95% confidence interval bounded by the inner lines and 95% prediction interval.

APPLICATIONS

Example from the Biomedical Literature

The following example is described in Glanz[2], pgs. 244-250, and is taken from a 1978 paper[13] in the journal Radiology.

The heart's ability to pump blood depends on the volume of the left ventricle at the beginning of ventricular contraction. One way to estimate left ventricular (LV) volume is to use angiography to inject into the heart a dye visible with x-rays. Two perpendicular x-ray views can be used to calculate LV volume. However, the measurements must first be calibrated by x-raying casts of ventricles from cadavers, for which actual LV volumes are known. Regressions of the LV volume estimated by

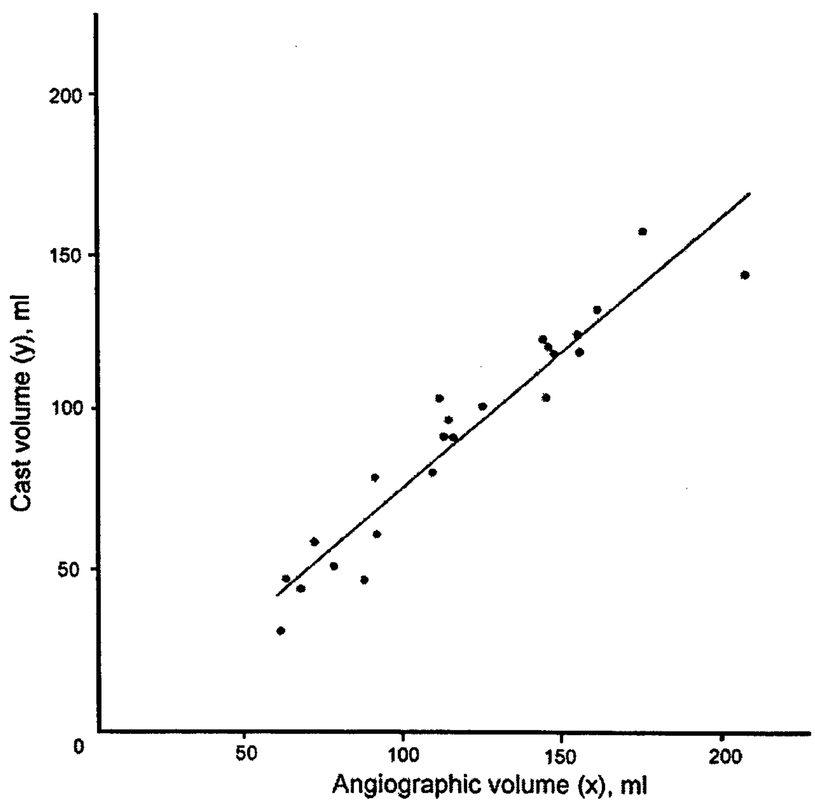


Figure 7. The linear regression of cast volume against angiographic volume.

angiography reveal a linear relationship with the actual cast volumes (Figure 7). Figure 8 has lines representing the "95% confidence interval for an additional observation *(in more modern terminology a 95% prediction interval*) of cast volume given angiographic volume. This is the confidence (*prediction*) interval that should be used to estimate true left ventricular volume from an angiogram to be 95% confident that the range includes the true volume." Reading from Figure 8, when a patient has an angiographic volume of

100 mL, we predict with 95% confidence that the true volume of the patient's left ventricle is between 60 and 105 mL. Use of the prediction interval thus results in a clinically useful estimate of LV volume.

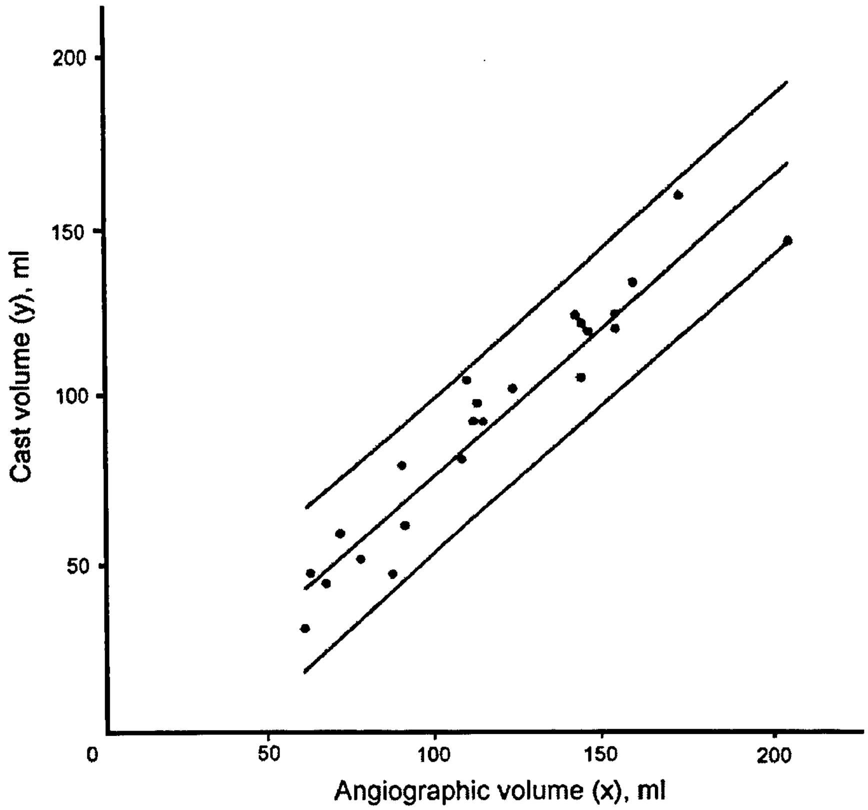


Figure 8. 95% prediction limits for cast volume as a function of angiographic volume.

Tool Life Example

The following example is from McClave and Benson, A First Course in Business Statistics[7]. The data was obtained from pg. 538, applied to problem 11.53b on pg. 564. We have supplemented this analysis to better illustrate the usefulness of prediction limits.

A manufacturer measured the effective life of a commonly used cutting tool. The manufacturer performed triplicate measurements at 5 cutting speeds: 30, 40, 50, 60, and 70 meters per minute. The data and linear regression of tool life against speed for the tool is shown in Figure 9. The figure also shows the 90% prediction intervals across the range of tested tool speeds. From this figure we see that at a cutting speed of 45 m/min, we predict with 90% confidence that the tool would have a minimum tool life of 3.4 hrs. The maximum life we would expect would be 5.6 hrs, with an average of 4.5 hrs.

This manufacturer has learned that it is cheaper to replace a tool before it becomes dull than to redo work marred by a dull tool. Therefore, the manufacturer wishes to replace the tools after a running time that is known to result in no more than 5% of the tools becoming dull. The manufacturer decided that although a zero rework rate was desired, up to a 5% rework rate could be tolerated without impacting production schedules or manufacturing costs.

On the other hand, there is an economic incentive in <u>not</u> replacing tools prematurely. Replacing tools every hour would certainly keep rework rates down, but the costs associated with frequently halting a production run, and of unnecessarily replacing tools, argue against being arbitrarily conservative. The best way to manage the competing requirements is to use statistically derived prediction limits, and to apply them to the matter of manufacturing process control.

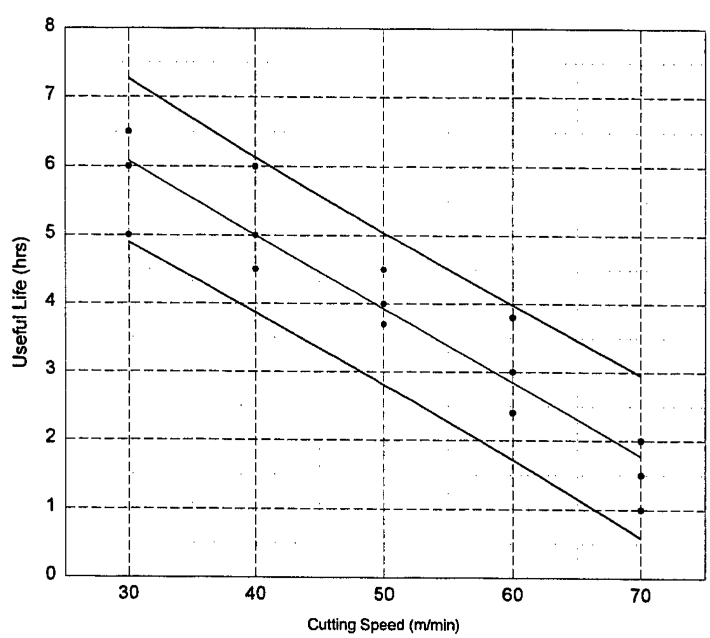


Figure 9. Tool life as a function of cutting speed.

Different cutting processes require different cutting speeds, so no single replacement time would suffice for all processes. We see from Figure 9 that for a process requiring a cutting speed of 45 m/min, the replacement time that would result in no more than 5% of the tools becoming dull is 3.4 hrs, the lower limit of the 90% prediction interval seen in Figure 9. This result comes from the fact that an estimated 90% of the tool lives fall between the two 90% prediction limits. That means that 5% of the tool lives will fall below the lower limit, and 5% will exceed the upper limit. It is the lower 5% figure that controls the replacement time.

This example is particularly relevant to the issue of canister durations. For one thing, the tool life data was obtained by replicate measurements at multiple running speeds. Canister durations are also made with five replicate runs at four or five temperatures ranging from 28° to 90° F or more. The variable of interest to the manufacturer is tool life, just as divers need to know canister life. The controlling statistic is the lower prediction limit for both the tool life example and for canister life. In both cases, the critical component is to be replaced before the probability of failure becomes too large. Also in each case there is a strong incentive for not being unduly conservative. In one case, overall manufacturing efficiency would suffer from over-conservatism. Likewise, in a military mission, over-conservatism of canister duration limits (limits that are too short) would impact operational efficiency.

Comparison of Current and pre-1991 Canister Limit Methods

In Figures 10 and 11 we have again sampled the Truth data set, this time to better illustrate the differences between the old and new method of deriving canister limits. As before, we have withdrawn 25 samples from a data set with a mean of 200 min. This time, however the data set has a standard deviation (σ) of 10 instead of 20. Again five data points were sampled at each of five temperatures spread between 32° and 90°F. In reality, there was no effect of temperature on the durations.

In the first round of sampling, upper panel of Figure 10, we see that the means of the samples at all five temperatures (filled circles) varied only slightly from the true population mean of 200 min. However, the canister limit defined by the sample mean minus one standard deviation (open circles) varied considerably, with a large dip at 70° F. Although we might have believed we were defining a canister limit with a 16% risk of breakthrough by using this method, the true risk varied from 3.5% to 29%, as indicated by the true risk lines running horizontally across the panel. In the lower panel, both the slightly curving fitted mean duration line and the lower 95% prediction limit closely approximate the true 50% and 2.5% risk lines, just as they should. That means that the prediction limit method of determining canister durations does a much better job of estimating true risk of canister breakthrough, and of identifying true temperature dependencies, than does the $\mu - \sigma$ method.

In the second round of sampling, Figure 11, the estimate of the mean durations varied about three times as much as in Figure 10. Because the means at both ends of the temperature scale happened to be relatively low, regression yielded a curvilinear prediction limit. The lower prediction limit designed to indicate a 2.5% risk, did so at the temperature extremes. However, it rose to as much as 6% risk at 60°F. Even then, the estimated risk was only off by 3.5%. By the $\mu - \sigma$ method, the actual risk exceeded the 16% target risk by as much as 16%, for a total of 32% risk. On an absolute risk scale, the prediction limit method yielded substantially reduced errors when by chance the data appeared to be skewed.

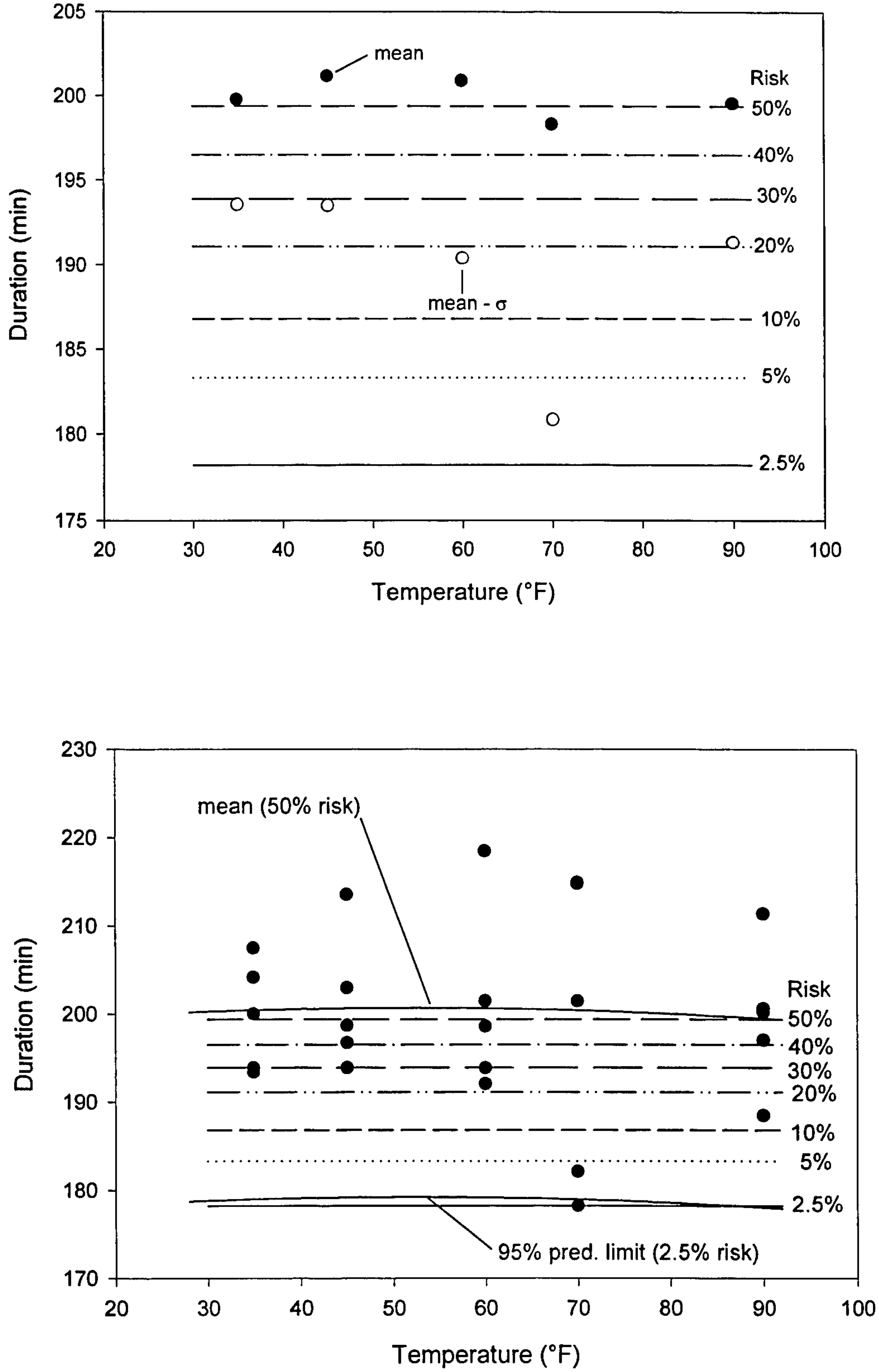


Figure 10. Pre-1991 method (upper panel) and new method (lower panel) of determining canister duration limits. Random data selection - no real temperature effect.

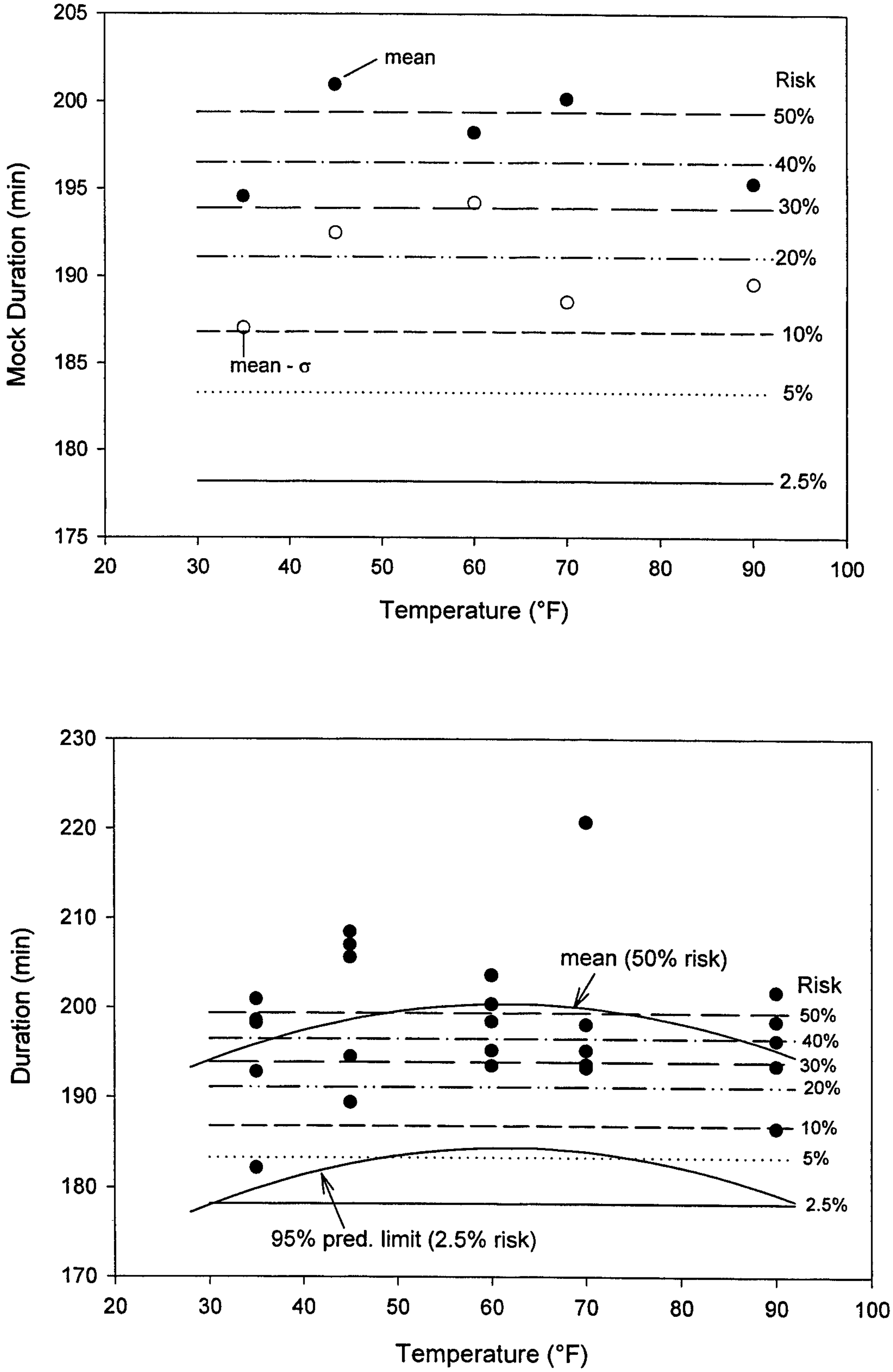


Figure 11. Old NEDU method (upper panel) and new method (lower panel) of determining canister duration limits. Second random data selection - no real temperature effect.

METHODS

Regression Model Development

The remainder of this report describes the process of generating models, fitting them to the data, and deriving predictions limits. It concludes with examples taken from two UBA.

We generate a mathematical model to fit to the experimental canister data in the following manner.

Assumptions:

1) The independent variable (X) is temperature.

2) The dependent variable (Y) is canister duration in minutes.

3) Y is a continuous function of X (dY/dX not equal to ∞). In essence, that means Y can be described by a simple equation.

4) Y is real at X > 0°K. Therefore one term (Y intercept) is required at a minimum.

5) Y = f(X). Therefore a second term must exist (e.g. Y = a + bX).

6) Y may be a curvilinear function of X: $Y = a + bX^c$

7) There is no *a priori* requirement for monotonicity. Therefore, there can be a 3rd term opposite in sign to the second term.

From these requirements, a group of potential models can be generated. Our purposes can be served by polynomial models, which are of the form:

$$Y = a + b \cdot X + c \cdot X^2 + d \cdot X^3 \ldots$$

Polynomials are a class of empirical models that vary in complexity from a single term to many terms. If the data set to be described is small, the number of terms that can be justified is also small. If too many terms are added to a model, we obtain a condition called over-parameterization, characterized by very large confidence intervals.

The optimal model for describing experimental data as a continuous function of temperature is one that has neither too few nor too many parameters. When dealing with a single class of models, i.e. polynomials, the optimal model typically yields the lowest standard error for the fit. The optimal model also results in the narrowest confidence and prediction intervals throughout the experimental temperature range, and therefore provides the longest canister duration limits.

Data Analysis and Curve Fitting Procedure

The following procedures will be illustrated with real, not simulated data, obtained during unmanned canister duration testing at NEDU.

Step 1

The analytical process for a series of canister runs begins with entering the water temperature and corresponding canister duration into separate columns in a spreadsheet. Using the graphing program Sigmaplot by Jandel Scientific (San Rafael, CA), we plot the raw data in a scatter plot as a function of temperature, as in the upper panel of Figure 12. If the canister durations appear linearly related to temperature, then Sigmaplot can be used to draw linear regression lines through the data. Frequently, the time required to reach both 0.5% and 1% CO_2 (closed and open circles in Figure 12, respectively) or higher may be plotted on the same graph. Another graph can be created to display summary statistics (i.e. mean and standard deviation) for each water temperature (Figure 12, lower panel.) The graph and spreadsheet can then be saved in a Jandel Workbook.

Step 2

The next step involves preparation for fitting the data in the spreadsheet to a mathematical equation or "model." While virtually any curve fitting software could be used, including Jandel's Sigmaplot, we find it convenient to use another Jandel product called TableCurve 2D.

The first step to using TableCurve 2D is to create a custom equation set, a process that only needs to be accomplished once. TableCurve can fit over 8,000 linear and nonlinear equations to data in one processing action. However, it is generally more useful, and much less risky technically, to restrict curve fitting to a small subset of equations. Based on the logic expressed in the previous section, we restrict our custom equation set to standard polynomial equations, and just to be complete, to Y-transformed polynomial equations. Table 3 lists 23 of those equations. Two of those equations, marked in gray, are of the greatest interest to us, again based on the preceding model generation logic. Those are the equation for a straight line, equation number 1, and the equation for a second order polynomial, equation 1003.

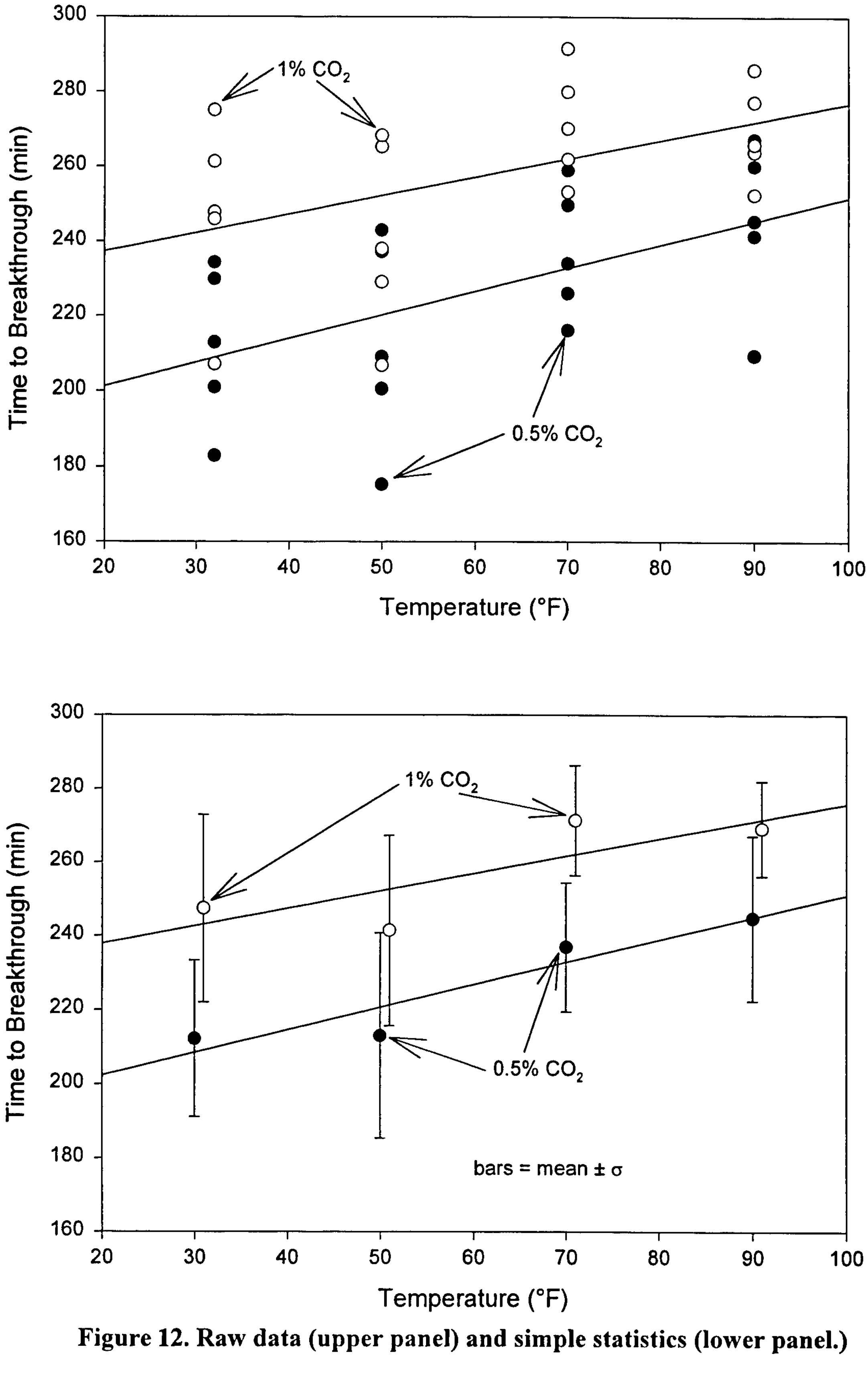


Figure 12. Raw data (upper panel) and simple statistics (lower panel.)

All of the equations following the third order polynomial, equation 2040, are Y-transformed equations. These transformations allow the linearization of an otherwise nonlinear equation, and are included only for completeness. We have yet to find a situation where they provide a better empirical fit to the data than the low order polynomials (the first three equations.)

Step 3

The next step is to import into TableCurve 2D the temperature and duration data from the Sigmaplot spreadsheet, with each variable being located within a separate column within the spreadsheet. From that point, the user need only select "Process Custom Equation Set" from the TableCurve Processing menu. At that point, TableCurve fits all of the custom set equations to the data, and then ranks them in order of best fit. Under "sort criteria" a number of choices are available. The one we find most useful is the F-statistic (Appendix B). That sorting preference can be made permanent by setting it in user preferences for TableCurve 2D.

Sorting by the F-statistic places at the top of the equation list the equation which best describes the data using the smallest number of parameters. Sorting in this manner minimizes the risk of over-fitting the data by the use of more parameters than are justified by the data (The TableCurve manual has a good discussion of the potential pitfalls involved in curve fitting and parameter estimation.) Table 4 is such a sorted list of fit equations. In that example, the best equation, that with the highest F-statistic, was equation 1, the equation for a straight line.

Step 4

The last step is to display the curve fitting results which include the line (which may be curved) representing the best estimate of the average or "mean" canister duration as a function of water temperature (Figure 13). It is also possible to plot the 95% confidence limits on the mean, thereby defining a region within which the true mean duration lies (with a 95% certainty).

Of greatest importance to establishing canister duration limits, is the plotting of 95% prediction limits (solid lines, Figure 13), two potentially curved lines which enclose 95% of the durations which are likely to occur in future dives. The lower of these limits is what NEDU uses as a canister duration limit. Ninety-seven and a half percent of canister durations are likely to fall above that line, and 2.5% of canister are likely to "break" before that dive time has elapsed, all else being equal.

Table 3. Custom TableCurve equation set for canister duration limits containing polynomials and Y-transformed polynomials.

Eqn. #	Equation
standard polynomials	
1	$y=a+bx$
1003	$y=a+bx+cx^2$
2040	$y=a+bx+cx^2+dx^3$
Y-transformed equations	
22	$\ln y=a+bx$
43	$y^{-1}=a+bx$
64	$y^{0.5}=a+bx$
85	$y^2=a+bx$
1213	$\ln y=a+bx+cx^2$
1423	$y^{-1}=a+bx+cx^2$
6101	$\ln y=a+bx+cx^2+dx^3$
6102	$\ln y=a+bx+cx^2+dx^3+ex^4$
6103	$\ln y=a+bx+cx^2+dx^3+ex^4+fx^5$
6111	$y^{-1}=a+bx+cx^2+dx^3$
6112	$y^{-1}=a+bx+cx^2+dx^3+ex^4$
6113	$y^{-1}=a+bx+cx^2+dx^3+ex^4+fx^5$
6121	$y^{0.5}=a+bx+cx^2$
6122	$y^{0.5}=a+bx+cx^2+dx^3$
6123	$y^{0.5}=a+bx+cx^2+dx^3+ex^4$
6124	$y^{0.5}=a+bx+cx^2+dx^3+ex^4+fx^5$
6131	$y^2=a+bx+cx^2$
6132	$y^2=a+bx+cx^2+dx^3$
6133	$y^2=a+bx+cx^2+dx^3+ex^4$
6134	$y^2=a+bx+cx^2+dx^3+ex^4+fx^5$

Table 4. Fit equations sorted by their F-statistic. The best fit is at the top.

Curvefit Rankings for 0.5% SEV CO_2

Rank	F-statistic	Eqn #	Equation
1	7.962883557	**1**	$y=a+bx$
2	7.8601110639	**64**	$y^{0.5}=a+bx$
3	7.5283512645	**22**	$\ln y=a+bx$
4	7.3357010958	**85**	$y^2=a+bx$
5	6.3801158367	**43**	$y^{-1}=a+bx$
6	3.8079206858	**1003**	$y=a+bx+cx^2$
7	3.742891358	**6121**	$y^{0.5}=a+bx+cx^2$
8	3.5727028219	**1213**	$\ln y=a+bx+cx^2$
9	3.5546937052	**6131**	$y^2=a+bx+cx^2$
10	3.0133372321	**1423**	$y^{-1}=a+bx+cx^2$
11	2.7408796673	**2040**	$y=a+bx+cx^2+dx^3$
12	2.6986270263	**6122**	$y^{0.5}=a+bx+cx^2+dx^3$
13	2.5801276758	**6101**	$\ln y=a+bx+cx^2+dx^3$
14	2.5622800357	**6132**	$y^2=a+bx+cx^2+dx^3$
15	2.1798855221	**6111**	$y^{-1}=a+bx+cx^2+dx^3$

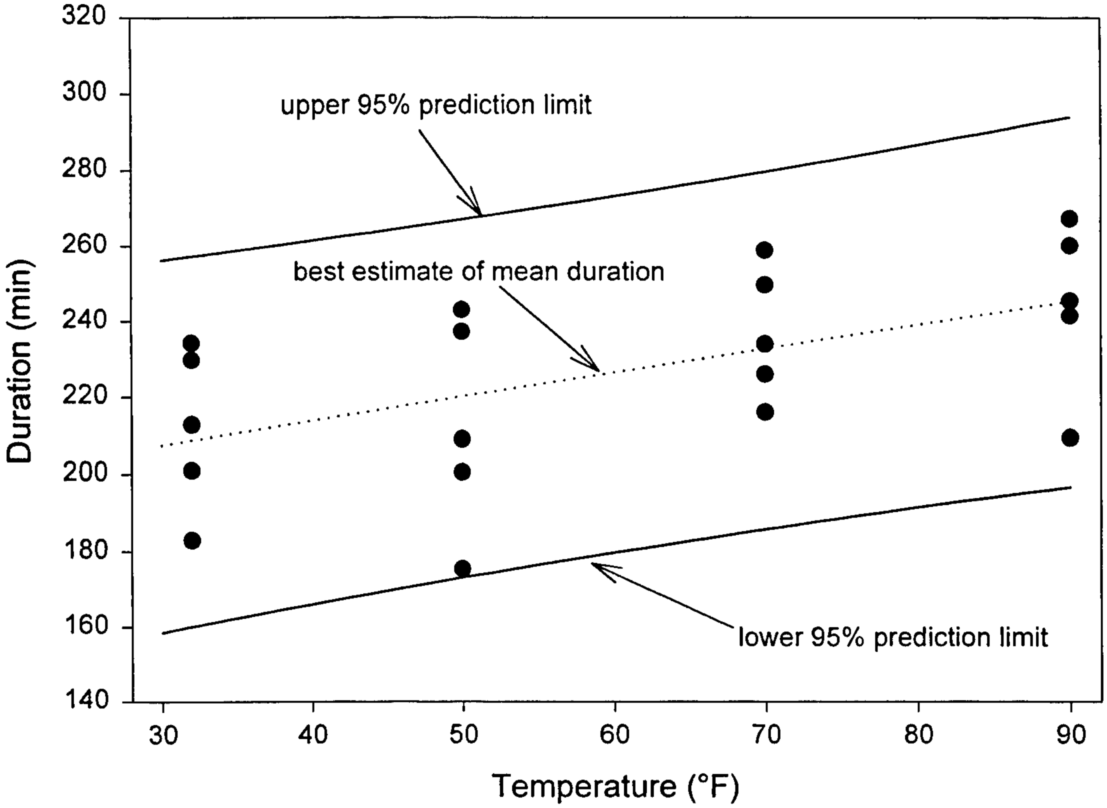
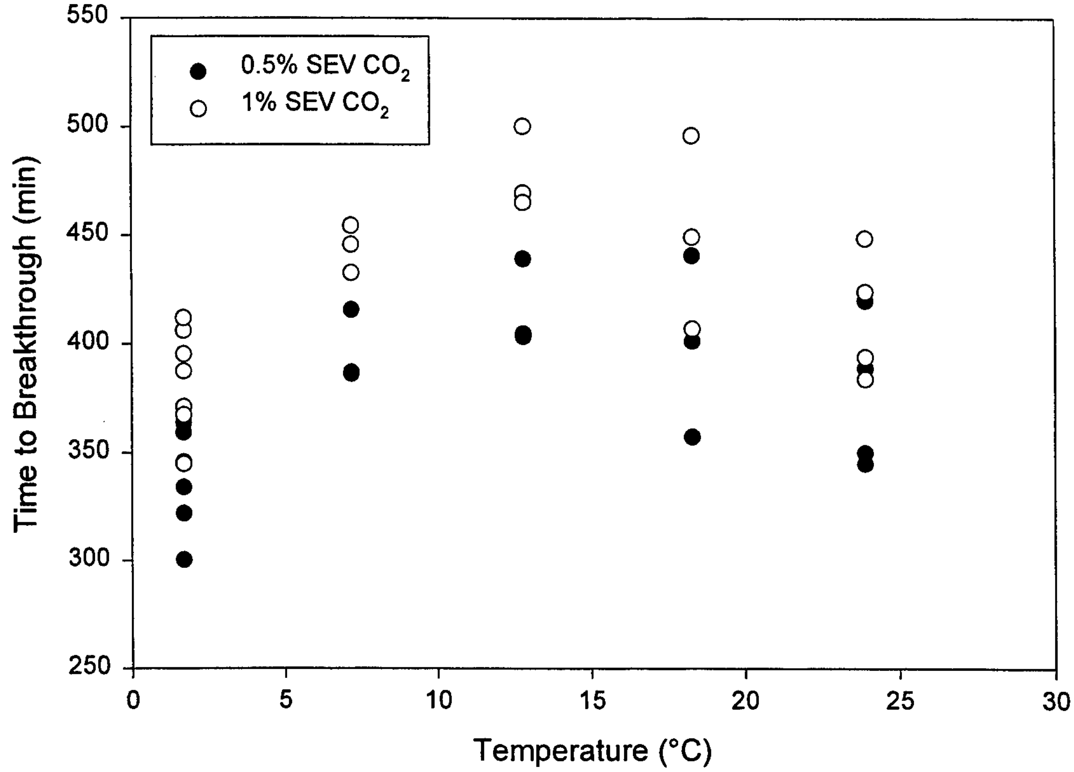


Figure 13. Best fit linear regression (Equation 1) and 95% prediction limits.

<u>Curvilinear Temperature Dependence</u>

In the next example based again on real, not simulated data, canister durations were curvilinearly related to temperature, with duration dropping at both ends of the temperature scale (Figure 14.) At the ends of the measured temperature range, data for the 0.5% and 1% CO_2 levels overlap.

Due to the pronounced curvilinearity in this data, the curvefit ranking favors the second order polynomial over other equations for both the 0.5% (Table 5) and 1% CO_2 (Table 6) measurement points. The results of the curvefitting are shown in Figure 15, the top panel showing results for the 0.5% CO_2 data and the bottom panel with 1% CO_2 data.



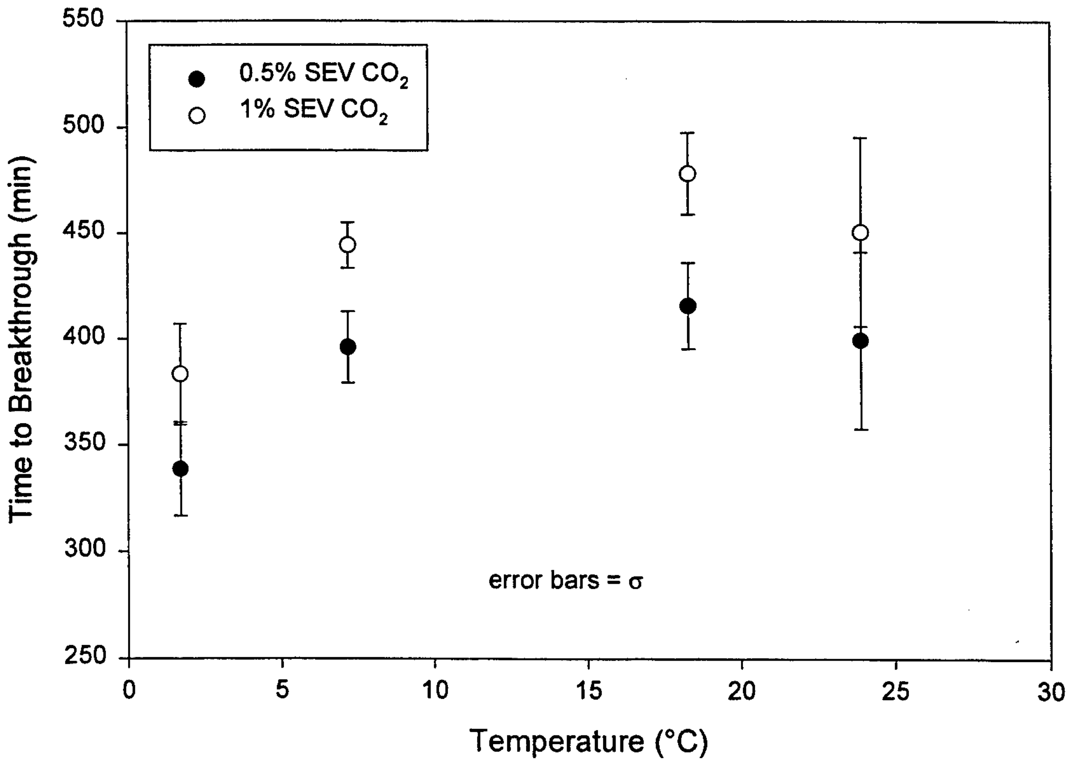


Figure 14. Canister duration data from a canister with a curvilinear temperature dependency. Upper panel is raw data for 0.5% and 1% SEV CO_2. The lower panel shows the mean for each temperature and %CO_2, with one standard deviation error bars.

Table 5. Curvefit rankings for Figure 14. Sorted by the F-statistic.

Curvefit Rankings for 0.5% SEV CO_2

Rank	F Statistic	Eqn #	Equation
1	12.186480729	**1003**	$y=a+bx+cx^2$
2	12.069139485	**6121**	$y^{0.5}=a+bx+cx^2$
3	12.052053317	**6131**	$y^2=a+bx+cx^2$
4	11.831418656	**1213**	$\ln y=a+bx+cx^2$
5	11.036941569	**1423**	$y^{-1}=a+bx+cx^2$
6	7.9973458536	**2040**	$y=a+bx+cx^2+dx^3$
7	7.9662598388	**6122**	$y^{0.5}=a+bx+cx^2+dx^3$
8	7.8565916688	**6101**	$\ln y=a+bx+cx^2+dx^3$
9	7.8212708657	**6132**	$y^2=a+bx+cx^2+dx^3$
10	7.4276516274	**6111**	$y^{-1}=a+bx+cx^2+dx^3$
11	5.6462788687	**6001**	$y=a+bx+cx^2+dx^3+ex^4$
12	5.6175850374	**6123**	$y^{0.5}=a+bx+cx^2+dx^3+ex^4$
13	5.5332519665	**6133**	$y^2=a+bx+cx^2+dx^3+ex^4$
14	5.5330958886	**6102**	$\ln y=a+bx+cx^2+dx^3+ex^4$
15	5.2177171699	**6112**	$y^{-1}=a+bx+cx^2+dx^3+ex^4$
16	4.5059832247	**1**	$y=a+bx$
17	4.3254555319	**85**	$y^2=a+bx$
18	4.2794315608	**64**	$y^{0.5}=a+bx$
19	4.2158882219	**6002**	$y=a+bx+cx^2+dx^3+ex^4+fx^5$
20	4.1944634946	**6124**	$y^{0.5}=a+bx+cx^2+dx^3+ex^4+fx^5$
21	4.1314948016	**6134**	$y^2=a+bx+cx^2+dx^3+ex^4+fx^5$
22	4.1313782635	**6103**	$\ln y=a+bx+cx^2+dx^3+ex^4+fx^5$
23	3.8958954868	**6113**	$y^{-1}=a+bx+cx^2+dx^3+ex^4+fx^5$
24	3.8566055685	**22**	$\ln y=a+bx$
25	3.2622944574	**6003**	$y=a+bx+cx^2+dx^3+ex^4+fx^5+gx^6$
26	2.5811560542	**6004**	$y=a+bx+cx^2+dx^3+ex^4+fx^5+gx^6+hx^7$
27	2.5425873307	**43**	$y^{-1}=a+bx$

Table 6. Curvefit rankings for Figure 14. Sorted by the F-statistic.

Curvefit Rankings for 1% SEV CO_2

Rank	F Statistic	Eqn #	Equation
1	17.837087478	**1003**	$y=a+bx+cx^2$
2	17.815211472	**6121**	$y^{0.5}=a+bx+cx^2$
3	17.683478283	**1213**	$\ln y=a+bx+cx^2$
4	17.547664572	**6131**	$y^2=a+bx+cx^2$
5	17.113176199	**1423**	$y^{-1}=a+bx+cx^2$
6	11.346228467	**6122**	$y^{0.5}=a+bx+cx^2+dx^3$
7	11.332728546	**2040**	$y=a+bx+cx^2+dx^3$
8	11.281373915	**6101**	$\ln y=a+bx+cx^2+dx^3$
9	11.068585548	**6132**	$y^2=a+bx+cx^2+dx^3$
10	10.936302935	**6111**	$y^{-1}=a+bx+cx^2+dx^3$
11	8.3077799934	**6001**	$y=a+bx+cx^2+dx^3+ex^4$
12	8.2779953127	**6123**	$y^{0.5}=a+bx+cx^2+dx^3+ex^4$
13	8.1901135997	**6102**	$\ln y=a+bx+cx^2+dx^3+ex^4$
14	8.1898252897	**6133**	$y^2=a+bx+cx^2+dx^3+ex^4$
15	7.8588054962	**6112**	$y^{-1}=a+bx+cx^2+dx^3+ex^4$
16	6.2031423951	**6002**	$y=a+bx+cx^2+dx^3+ex^4+fx^5$
17	6.1809031668	**6124**	$y^{0.5}=a+bx+cx^2+dx^3+ex^4+fx^5$
18	6.1152848211	**6103**	$\ln y=a+bx+cx^2+dx^3+ex^4+fx^5$
19	6.1150695497	**6134**	$y^2=a+bx+cx^2+dx^3+ex^4+fx^5$
20	5.8679081038	**6113**	$y^{-1}=a+bx+cx^2+dx^3+ex^4+fx^5$
21	4.8000506628	**6003**	$y=a+bx+cx^2+dx^3+ex^4+fx^5+gx^6$
22	3.7978422827	**6004**	$y=a+bx+cx^2+dx^3+ex^4+fx^5+gx^6+hx^7$
24	2.9341911296	**1**	$y=a+bx$
25	2.7486850719	**64**	$y^{0.5}=a+bx$
26	2.7134568482	**85**	$y^2=a+bx$
28	2.3695816642	**22**	$\ln y=a+bx$
30	1.1246693055	**43**	$y^{-1}=a+bx$

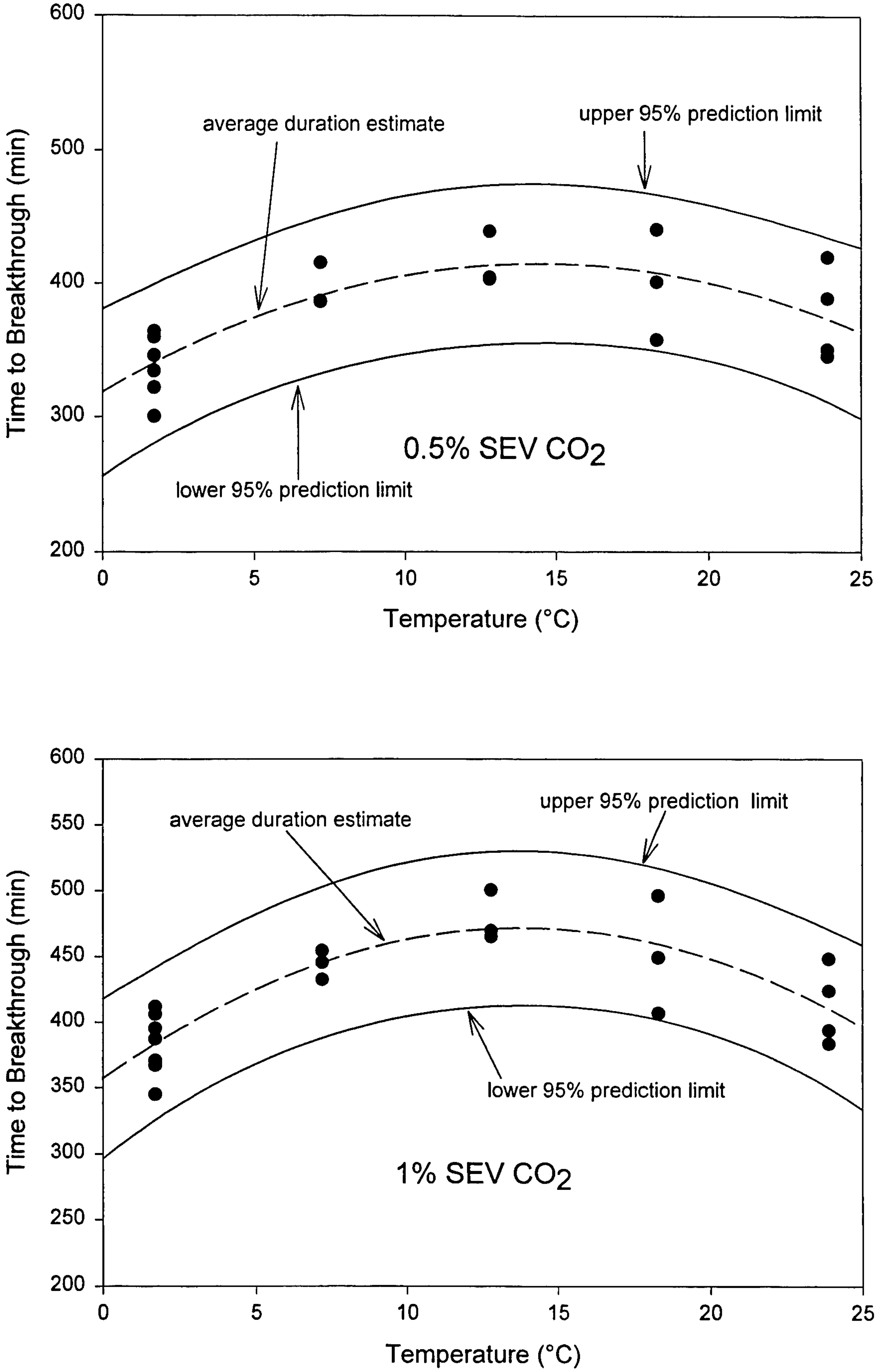


Figure 15. Best curvefits for 0.5% and 1% CO_2 data.

There is a penalty to pay if one ignores the curvilinearity of the canister duration data and base canister limits instead on a linear equation (Equation 1). As shown in Figure 16, that penalty is an overly conservative canister limit throughout the range of the majority of the data. In other words, our modeling effort is repaid by the maximum canister duration allowed by the data.

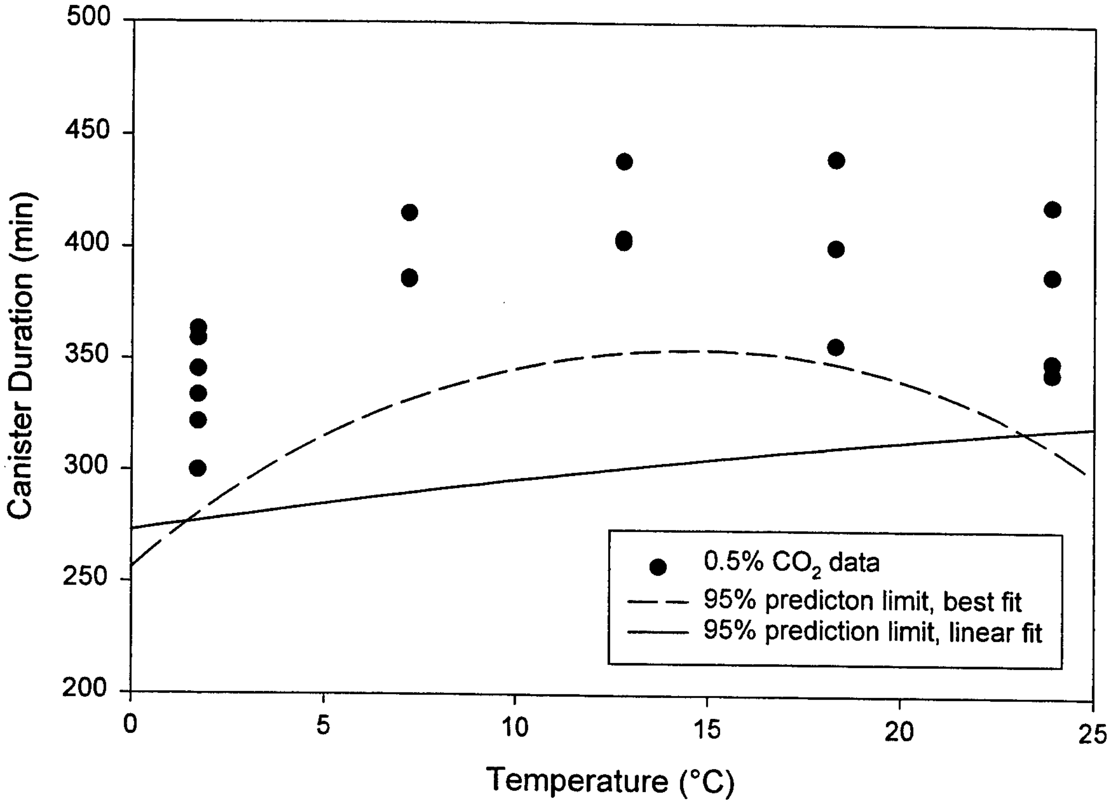


Figure 16. Comparison of the lower 95% prediction limits resulting from a linear model (solid line, F = 4.5) and the best fit model, a second-order polynomial (dashed line, F = 12.2).

CONCLUSIONS

Regression-based prediction limits based on empirical models of CO_2 canister performance provide an enhancement over the older methods of determining canister duration limits. They make maximum use of the data, thereby providing the longest possible canister duration for a given level of canister breakthrough risk.

THIS PAGE BLANK

REFERENCES

1. L. Thomas, "Biostatistics in Medicine", *Science*, 198:675, 1977.

2. S.A. Glanz, *A Primer of Biostatistics*, 3rd ed., McGraw Hill, Inc., New York, pg. 248-249, 1992.

3. R.D. Gafford, M.W. Greene, "Monitoring Carbon Dioxide Partial Pressures in Semi-Closed Underwater Breathing Apparatus", In: *Equipment for the Working Diver, 1970 Symposium. Symposium Proceedings*. Marine Technology Society, Washington D.C., pg. 283-297, 1970.

4. C. Lentner, ed. *Geigy Scientific Tables*, CIBA-Geigy, New Jersey, pg. 230, 1981.

5. C. F. Consolazio, R.E. Johnson, L.J. Pecora, *Physiological Measurements of Metabolic Functions in Man*. McGraw-Hill, New York. pg. 53, 1963.

6. R.R. Sokal, F.J. Rohlf, *Biometry, the Principles and Practice of Statistics in Biological Research*, 3rd ed, W.H. Freeman and Co., New York, pg. 472-475, 1995.

7. J.T. McClave, P.G. Benson, *A First Course in Business Statistics*, 5th ed. MacMillan Publishing Company, New York, pg. 558-562, 1992.

8. J. Neter, W. Wasserman, M.H. Kutner, *Applied Linear Statistical Models, Regression, Analysis of Variance, and Experimental Designs*, 3rd ed, Richard D. Irwin Inc., Homewood, IL, pg. 79-85, 1990.

9. J. Cohen, P. Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 2nd ed. Lawrence Erlbaum Associates, New Jersey, pg. 63-64, 1983.

10. N. Draper, H. Smith, *Applied Regression Analysis*, 2nd ed. John Wiley & Sons, New York, pg. 30-31, 1981.

11. J. Clarke, L. Crepeau, K. Russell, *MK 16 Canister Limits for SDV Operations*. Navy Experimental Diving Unit, NEDU TR 2-93, pg. 4-8, 1993.

12. J. Clarke, L. Crepeau. *Recommended Canister Limits for the Draeger LAR V/MK 25 UBA Using 408 L-grade and 812 D-Grade Sofnolime*. Navy Experimental Diving Unit, NEDU TR 9-97, pg. 6-8, 1993.

13. M.J. Lipton, T.T. Hayashi, D. Boyd and E. Carlson. "Measurement of left ventricular volumes by computed tomography", *Radiology*, 127:419-423, 1978.

14. H. Scheffé, *The Analysis of Variance*, John Wiley & Sons, New York, pg. 331-369, 1959.

THIS PAGE BLANK

Appendix A: Variable Risk Prediction Limits

The TableCurve software provides 95%, 90%, and 50% prediction limits (with the 50% limit being based on the estimated mean duration for any given temperature). However, the process of generating those limits provides intermediate results that can be used to generate other prediction limits. For example, Figure A1 shows prediction limits varying in risk from 2.5% to 50%. The 10%-30% values were obtained from an additional analysis described below.

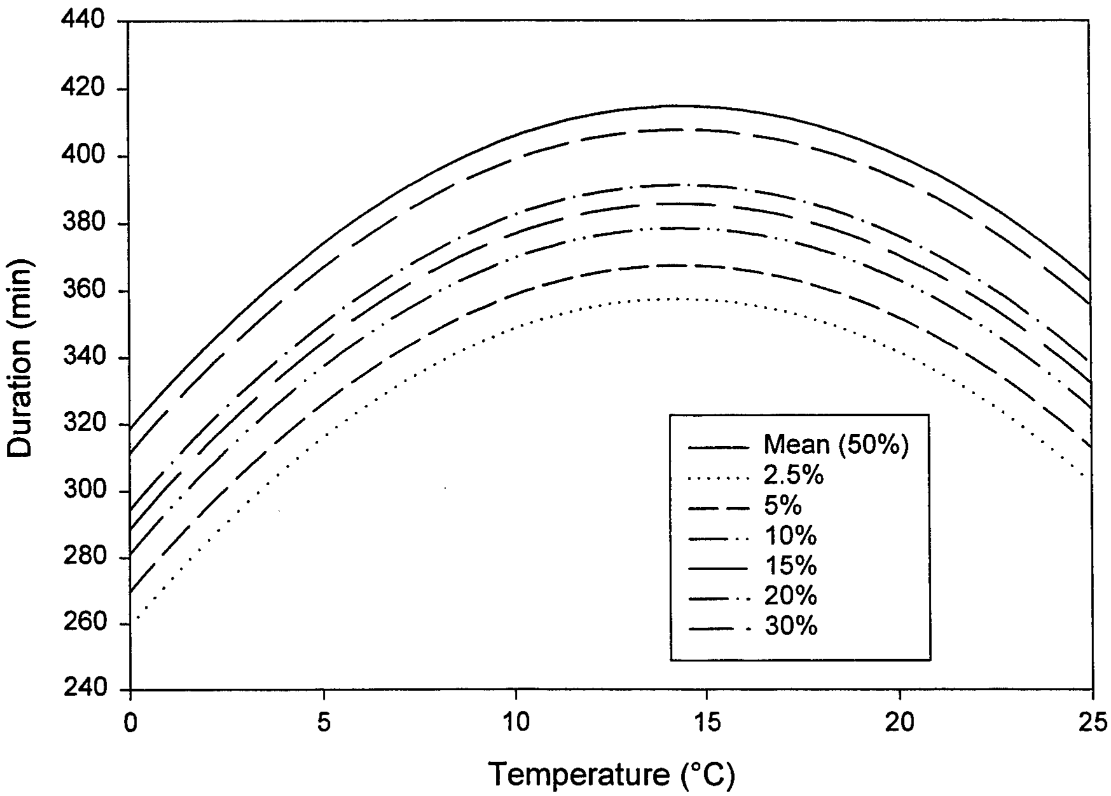


Figure A1. Prediction limits for risk levels ranging from 2.5% to 50%.

TableCurve outputs a numerical summary as shown in Table A1. In that Table we highlighted results that are essential to the follow-on analysis; namely, Fit Standard Error (s_e), degrees of freedom for error (DOF), Xmean ($\overline{X}$), and Xstd or standard deviation of X values (s_x).

From these values we find the standard error ($\hat{s}$) of a predicted $\overline{Y}_i$ for a given X_i by the following equation:

$$\hat{s}_{\overline{Y}_i} = s_e \cdot \sqrt{(1+\frac{1}{n}) + \frac{(X_i - \overline{X})^2}{(n-1) \cdot s_x^{\ 2}}}$$

where n is the number of data points.

Finally, the upper and lower prediction limits (L) are found as follows:

$$L = \hat{Y}_i \pm (t_{\%risk} \cdot \hat{s}_{\bar{Y}_i})$$

where the $t_{\%risk}$ is the value found from a table of the critical values of Student's t-distribution, found in any statistics book, with degrees of freedom (DOF) appropriate for the problem, and the desired percentage risk.

The degrees of freedom given in the TableCurve numerical summary is nothing more than the number of data points minus the number of fit parameters. For the example of Figures 15 and 16, there were 20 data points and three parameters (a-c) in the best fit model (the top ranked polynomial equation #1003 in Table 6). Thus the appropriate degrees of freedom is 17.

To find the proper t value, we take the allowed risk, say 2.5%, and express it as a decimal (0.025). Then we look up in the t table the appropriate value of t for 17 degrees of freedom (v, usually given in rows) and a one-tailed critical area equal to 0.025 (usually found in columns). For this particular case, $t_{0.05(17)} = 2.1098$. If the t table happens to be for two tails of the t-distribution, then the allowed risk should be multiplied by two, and the proper t would be found under the column for $t_{0.05(17)}$.

NEDU uses in-house developed software that computes upper and lower prediction limits for any X_i. In addition to the above inputs, the only remaining input is a file containing all of the X_I and $\hat{Y}_i$ generated after the TableCurve fitting procedure. That data file is generated by "Evaluating" (one of the TableCurve options) the best model between the upper and lower bounds of the existing data set. By evaluating the best fit equation for many values of X, means and prediction limits can be found for X values falling between those in the actual data set.

Table A1. Numeric Output from TableCurve.

Rank 1 Eqn 1003 $y=a+bx+cx^2$

r^2 Coef Det	DF Adj r^2	Fit Std Err	F-value
0.5891036222	0.5120605514	26.436260097	12.186480729

Parm	Value	Std Error	t-value	95% Confidence Limits		P>|t|
a	318.5653904	12.98603957	24.53137376	291.1672418	345.963539	0.00000
b	13.34412691	2.916198232	4.575864139	7.19148645	19.49676737	0.00027
c	-0.46290207	0.115402947	-4.01118062	-0.706381	-0.21942313	0.00091

Area Xmin-Xmax	Area Precision		
8758.2763501	0		
Function min	X-Value	Function max	X-Value
339.91265596	1.7000031239	414.73353073	14.413553014
1st Deriv min	X-Value	1st Deriv max	X-Value
-8.7825918	23.9	11.770256997	1.7000031239
2nd Deriv min	X-Value	2nd Deriv max	X-Value
-0.925804179	19.989509283	-0.925804081	15.8610701

Soln Vector	Covar Matrix		
Direct	LUDecomp		
r^2 Coef Det	DF Adj r^2	Fit Std Err	r^2 Attainable
0.5891036222	0.5120605514	26.436260097	0.6009058424

Source	Sum of Squares	DOF	Mean Square	F Statistic	P>F
Regr	17033.674	2	8516.8371	12.1865	0.00052
Error	11880.889	17	698.87585		
Total	28914.564	19			
Lack Fit	341.25604	2	170.62802	0.221794	0.80366
Pure Err	11539.633	15	769.30889		

Description: D:\UBA\LARV\MK25\closed-circuitUBA

X Variable: tmp

Xmin:	1.7	Xmax:	23.9	Xrange:	22.2
Xmean:	11.12	Xstd:	8.8368010406	Xmedian:	10
X@Ymin:	1.7	X@Ymax:	18.3	X@Yrange:	16.6

Y Variable: 0.5%*.72

Ymin:	300.24	Ymax:	440.64	Yrange:	140.4
Ymean:	375.372	Ystd:	39.010500287	Ymedian:	374.76
Y@Xmin:	363.6	Y@Xmax:	349.92	Y@Xrange:	13.68

THIS PAGE BLANK

Appendix B: The F-statistic

The next two paragraphs are quoted from the Jandel TableCurve manual, pgs 8-30 and 8-31.

"The F-statistic is a measure of the extent to which the given equation represents the data. If an additional parameter makes a statistically significant contribution to a model, the F-statistic increases. Otherwise, a decrease occurs. The higher the F-statistic, the more efficiently a given equation models the data.

In a parametric model, the main interest is in the value of the parameters or coefficients. In such cases, the number of parameters is very important and their values are usually the object of the fitting. If you are fitting non-linear equations, you will almost always wish to use the F-statistic option as the sort criteria. Even when fitting linear equations for an approximating function, the F-statistic option may be an effective way of ordering the equations so that the simpler and yet effective equations find their way nearer the top of the equation list."

The F-statistic is most often encountered in discussions of analysis of variance (ANOVA). It is actually a ratio, and thus it is often called the F-ratio. It is defined as:

$$F = \frac{MSR}{MSE} = \frac{\dfrac{SSM - SSE}{m-1}}{\dfrac{SSE}{DOF}}$$

where MSR is the so-called "mean square regression" and MSE is the "mean square error". SSM is the sum of squares about the mean, and SSE is the sum of squares due to error, DOF is degrees of freedom, and m is the number of coefficients fitted.

The SSE measures the discrepancy between a predicted canister duration and a measured duration. Thus it reflects the amount of error in the prediction.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

The SSM describes how much the y values (canister duration) vary around their mean value.

$$SSM = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The F-statistic becomes large when the amount of variation around the mean y is large compared to the differences between predicted and actual y's. In general. the larger the F-statistic, the better the fit of the regression curve to the data.

The coefficient of determination (r^2), commonly encountered in regression, shares some attributes in common with the F-statistic.

$$r^2 = 1 - \frac{SSE}{SSM}$$

The degree of freedom adjusted r^2 is even more similar to the F-statistic:

$$r^2 = 1 - \frac{SSE \cdot (n-1)}{SSM \cdot (DOF-1)}$$

Nevertheless, only the F-statistic takes into account the number of parameters (m) fit to the data. All else being equal, a model which uses two parameters will provide a larger F-statistic than one using three parameters. The F-statistic thus guards against over-fitting the data; using more parameters than are necessary to describe the data. Using parameters that do not produce a significant improvement of the model fit only serve to reduce the F-statistic.

Like all of the statistical methods discussed in this report, the F-statistic requires some assumptions about the data. The effect of departures from the underlying assumptions are rigorously described in Scheffe's *The Analysis of Variance*[14].